

د. ثروت محمد عبد المنعم

y

70

60

50

40

30

20

10

# الانحدار

20

40

60



مكتبة الأنجلو المصرية



# الإنحدار

تأليف دكتورة  
شروت محمد عبد المنعم  
أستاذ مشارك بكلية العلوم بالدمام  
قسم الرياضيات



مكتبة الأنجلو المصرية

١٦٥ ش محمد فريد - القاهرة

اسم الكتاب : الإنحدار

اسم المؤلف : دكتورة ثروت محمد عبد المنعم

اسم الناشر: مكتبة الأنجلو المصرية

اسم الطابع: مطبعة محمد عبد الكريم حسان

رقم الإيداع : ١٣٢٣٦ / ٢٠٠٥

الترقيم الدولي : 3 - 2149 - 05 - 977 - I.S.B.N



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

قال تعالى : ﴿ إِنَّ الَّذِينَ آمَنُوا  
وَعَمِلُوا الصَّالِحَاتِ إِنَّا لَا نَضِيعُ  
أَجْرَ مَنْ أَحْسَنَ عَمَلًا ﴾

صورة الكهف - الآية ٣٠



## الإهداء

إلى أختي في الله ..... د . فوزية البراهيم  
التي تتصف بالخلق الكريم والتدين وصفاء القلب والصدق  
والتي تعتبر بعض خصال رسول الله ﷺ



### شكر وتقدير

أقدم وافر شكري إلى طابعتي وأولادي بالفرقة الرابعة "رياضيات" لعام ١٤٢٤ هجري وهن كالتالي على حسب الترتيب الأبجدي:

١. بنا الشهراني
٢. تهاني الخضير
٣. فاطمة الشهري
٤. فاطمة ابالحارث
٥. مريم الشهاب
٦. منال المريحل
٧. نسرين آل الحارث
٨. نعمة المغربي
٩. نوف العنزي

على ما بذلته من جهد في المشاركة في الكتابة وتنفيذ البرامج على الحاسب الآلي.



### شكر و تقدير

أقدم وافر شكري إلى طالباتي وأولادي بالفرقة الرابعة "رياضيات" لعام ١٤٢٦ هجري ومن كالتالي على حسب للترتيب الأبجدي:

١. أمل الدورره
٢. تغريد ترحيب المطيري
٣. ثريا سلمان البحراني
٤. حصه ناصر الدوسري
٥. دينا يوسف الحمود
٦. ريم مغرم الشهري
٧. زهراء الخضراوي
٨. زينب أبو عبدالله
٩. سارة الزرمان
١٠. هبه علي الفاوي

على ما بذلته من جهد في المشاركة في الكتابة وتنفيذ البرامج على الحاسب الآلي.





## تقديم

يعتبر تحليل الانحدار من الموضوعات الهامة التي لا غنى عنها للباحثين في المجالات العلمية المختلفة ومن ثم فإنه يتعين على المهتمين ببناء النماذج الاقتصادية الإلمام الكافي بالفروض التي يجب توافرها في حالة تقدير نماذج الانحدار الخطية باستخدام طريقة المربعات الصغرى حيث أن إغفال فرض أو أكثر من هذه الفروض يترتب عليه أثار خطيرة فيما يتعلق بعملية التقدير للمعالم المراد اختبار معنوياتها . . ومن ناحية أخرى فعندما لا يكون النموذج المراد تقديره خطياً فإننا نكون بحاجة إلى التعرف على الطرق المختلفة لتقدير معالم العلاقات الغير الخطية وهذا الكتاب الذي تقدمه الدكتورة / ثروت محمد عبد المنعم يعتبر محاولة جيدة منها لتجميع كل ما هو متعلق بتحليل الانحدار الخطي سواء البسيط أو المتعدد أو الانحدار غير الخطي وكذلك مشاكل القياس المترتبة على إغفال فرض أو أكثر من فروض الانحدار الخطي في مرجع واحد ليقتابل احتياجات الدارسين المهتمين ببناء نماذج الانحدار وتحليلها . ومن الجدير بالذكر أن المادة العلمية التي يتضمنها هذا الكتاب تعتبر أساسيات فسي أدبيات الاقتصاد القياسي والذي يتضمن ضمن مستوياته هذه الأبواب بالإضافة إلى أبواب أخرى مختلفة ومتعددة تغطي كافة أنواع النماذج الاقتصادية سواء نماذج المعادلة الواحدة أو نماذج المعادلات الآتية إلى غير ذلك من الموضوعات ذات الصلة بتحليل الانحدار .

د.أ/ محمد عبد السميع عناني

الأستاذ بقسم الإحصاء والرياضة والتأمين

كلية التجارة - جامعة الزقازيق



## بسم الله الرحمن الرحيم

### تمهيد

الحمد لله رب العالمين والصلاة والسلام على أشرف المرسلين محمد وعلى آله وصحبه أجمعين. أما بعد، فالحمد لله الذي هدانا وما كنا لنهتدي لولا أن هدانا الله الذي أنعم علي بكتابة هذا الكتاب تلبية لنداء التعريب الذي يتناهى الكثير من العلماء والمتقنين.

تسمى أي طريقة لتوفيق معادلات لبيانات بالانحدار، تستخدم تلك المعادلات لغرضين على الأقل: عمل التنبؤات والحكم على قوة العلاقات. ولأن طرق الانحدار تمثدا بالكيفية التي يتأثر بها متغير ما بمتغيرات أخرى فإنها أصبحت ضرورية في معظم مجالات الدراسة التي تشتمل على العلوم الحيوية ، الفيزيائية ، العلوم الاجتماعية ، الصناعة ، الإقتصاد ، الطب ... الخ.

هذا الكتاب يصلح كمقرر لطلاب كليات العلوم ، كما يصلح لأن يكون مقرراً لطلاب الدراسات العليا في جميع مجالات البحث العلمي. هذا ويمكن لطلاب الدراسات العليا في المجالات التطبيقية مثل الزراعة والطب والهندسة التركيز على الجانب التطبيقي من هذا الكتاب وتتبع حل الأمثلة. يصلح هذا الكتاب أيضاً لأن يكون مرجعاً لأي باحث مع استشارة المتخصصين في الإحصاء وذلك لاختيار النموذج المناسب لتحليل البيانات، هذا ويمكن الاستعانة ببرامج الحاسب الآلي الخاصة بالانحدار وذلك لتنفيذ العمليات الحسابية مثل برامج SAS، SPSS، MINITAB. ويفضل إجابة أكثر من برنامج حتى يمكن الاستفادة من إمكانيات كل برنامج.

وفي وضع هذا الكتاب استعنت بكثير من المراجع العربية والأجنبية كما استعنت بخبرتي في تدريس هذا المقرر لطلاب الدراسات العليا في مرحلة الماجستير والدكتوراه وكما استعنت بخبرتي في الاستشارات الإحصائية في مرحلة الماجستير والدكتوراه ، ولذلك اشتمل الكتاب على العديد من الأمثلة في كافة المجالات. وقد استعنت بمجموعة من طالبات كلية العلوم للبنات بالدمام دفعة ١٤٢٤هـ ودفعة ١٤٢٦هـ قسم الرياضيات خلال بحث البكالوريوس لتنفيذ برامج Mathematica والتي أعدتها لمساعدتي في العمليات الحسابية و الكتابة الأولية لبعض فصول الكتاب، أيضاً استعنت ببرنامج SPSS وبرنامج Statistica لتنفيذ بعض الرسوم البيانية.

يحتاج الدارس لهذا الكتاب إلى معرفة أساسيات الإحصاء الرياضي و  
الإلمام بالجبر الخطي والتفاضل والتكامل حتى يستطيع أن يفهم كل محتويات هذا  
الكتاب.

يحتوي هذا الكتاب على عشرة فصول، يقدم الفصل الأول الانحدار الخطي  
البسيط، أما الفصل الثاني فيهتم بمخالفات فروض نموذج الانحدار الخطي البسيط  
وكيفية اكتشافها وتصحيحها، بينما يهتم الفصل الثالث بالانحدار الخطي المتعدد،  
ويتطرق الفصل الرابع إلى المخالفات وللخلل في فروض التحليل لنموذج الانحدار  
الخطي المتعدد وكيفية اكتشافها وتصحيحها، أما الفصل الخامس فيقدم اختبار  
أفضل نموذج لانحدار، كما يقدم الفصل السادس نماذج انحدارات كثيرات الحدود،  
ويهتم الفصل السابع بالمتغيرات الصورية كما يتطرق الفصل الثامن للارتباط  
الذاتي للأخطاء ويقدم الفصل التاسع الارتباط الخطي المتعدد، وأخيراً يتطرق  
الفصل العاشر إلى النماذج الغير خطية. هذا وتعتبر الفصول الأربعة الأولى هي  
الأساس الذي يجب على القارئ للتركيز عليه حتى يتمكن من فهم بقية الفصول  
الكتاب والتي تعتبر مواضيع متقدمة في الانحدار.

واسأل الله أن أكون قد وفقت في هذا المجهود المتواضع خدمة لقضايا  
البحث العلمي في وطننا العربي.

كما أتوجه بالشكر إلى دار للنشر التي أتاحت لي الفرصة لنشر هذا العمل  
العلمي.  
وبإني أرحب بكل نقد بناء يهدف إلى الأفضل، وما للكمال إلا الله وحده.

والله ولي التوفيق

## المحتويات

تقديم

تمهيد

### الفصل الأول : الإحدار الخطي البسيط

٢	(١-١) مفاهيم أساسية
٣	(٢-١) العلاقة بين متغير مستقل ومتغير تابع
٣	(١-٢-١) العلاقة الذاتية
٥	(٢-٢-١) العلاقة الإحصائية
٨	(٣-١) مقدمة في الإحدار الخطي البسيط
١٠	(٤-١) نموذج الإحدار الخطي البسيط
١٢	(٥-١) فروض لنموذج الإحدار الخطي البسيط
١٣	(٦-١) طريقة المربعات الصغرى
٢٦	(٧-١) خواص مقدرات المربعات الصغرى
٣٤	(٨-١) صيغة بديلة للنموذج
٣٨	(٩-١) تقدير $\sigma^2$
٤٤	(١٠-١) استدلالات لفحص معاملات الإحدار
٥٤	(١١-١) التنبؤ
٦٥	(١٢-١) أسلوب تحليل الإحدار
٧٣	(١٣-١) محامل التحديد
٧٨	(١٤-١) الإحدار خلال نقطة الأصل
٨٦	(١٥-١) الاستدلال آليا لمعامل النموذج
٩٦	(١٦-١) التقدير آليا لمتوسط الاستجابة
٩٨	(١٧-١) التنبؤ لعدد $m$ من مشاهدات جديدة
٩٩	(١٨-١) التقدير باستخدام الإمكان الأعظم
١٠١	(١٩-١) الارتباط

### الفصل الثاني : مغالطات فروض نموذج الإحدار الخطي البسيط وكيفية اكتشافها وتصحيحها

١١٣	(١-٢) مقدمة
١١٣	(٢-٢) تحليل البواقي
١١٤	(١-٢-٢) خواص البواقي
١١٨	(٢-٢-٢) رسوم البواقي
١٢٧	(٣-٢-٢) رسوم بواقي أخرى لاختبار الاعتدال
١٣٥	(٤-٢-٢) اختبار لنقص الاعتدال
١٣٨	(٣-٢) اختبار خطية الإحدار
١٥٣	(٤-٢) تحويلات إلى لخط المستقيم

## (٥.٢) اكتشاف وتصحيح عدم ثبات التباين

١٧٣	مقدمة (١.٥.٢)
١٧٣	طرق تحليلية لاكتشاف عدم ثبات التباين (٢.٥.٢)
١٧٤	تصحيح عدم ثبات التباين (٣.٥.٢)
١٨١	اختيار التحويلات (٦.٢)
٢١٩	تحويل قيم المتغير التابع (١.٦.٢)
٢١٩	طرق بيانية لتحويل قيم المتغير التابع أو قيمة للمتغير المستقل (٢.٦.٢)
٢٢٣	تحويل قيم المتغير المستقل (٣.٦.٢)
٢٢٦	وجود مشاهدة واحدة أو قليل من المشاهدات المتطرفة (٧.٢)
٢٢٩	

## الفصل الثالث : الإحدار الخطي المتعدد

٢٣٦	ملزمة (١.٣)
٢٣٨	تقدير المعلم (٢.٣)
٢٤٢	تقدير المعلم باستخدام المصفوفات (٣.٣)
٢٤٤	الإحدار البسيط في صيغة مصفوفة (٤.٣)
٢٤٦	فروض جاوس - ماركوف (٥.٣)
٢٤٧	خواص مقدرات المربعات الصغرى (٦.٣)
٢٥٠	خواص البواقي (٧.٣)
٢٥٥	صيغة أخرى للحصول على تقديرات المربعات الصغرى لمعلم نموذج الإحدار الخطي المتعدد (٨.٣)
٢٥٦	تقدير $\sigma^2$ (٩.٣)
٢٦١	فترات ثقة في الإحدار المتعدد (١٠.٣)
٢٦١	فترات ثقة لمعاملات الإحدار (١-١٠.٣)
٢٦٢	فترة ثقة لمتوسط الإستجابة (٢-١٠.٣)
٢٦٥	فترة ثقة لمشاهدة مستقبلية (٣-١٠.٣)
٢٦٧	فترة ثقة لدالة خطية لمعادلات إحدار (٤-١٠.٣)
٢٦٨	تقديرات أو تنبؤات خارج مجال النموذج (١١)
٢٦٩	اختبارات الفروض (١٢-٢١)
٢٦٩	اختبار يخص جميع معاملات الإحدار الجزئية (١-١٢.٣)
٢٧٣	معلم للتحديد المتعدد (٢-١٢.٣)
٢٧٥	اختبارات تخص كل معلم الإحدار (٣-١٢.٣)
٢٧٧	طريقة مجاميع المربعات الإضافية (٤-١٢.٣)

٢٨٥	(٥-١٢-٣) اختبار فرضية حول أهمية تعكس المتغيرات
٢٨٨	(٦-١٢-٣) الحالة الخاصة لأصدة متعلدة في المصفوفة $X$
٢٩٤	(٧-١٢-٣) اختبار الفرض الخطي $TB = 0$
٣٠٤	(١٢-٣) معاملات الإحدار القياسية
٣٠٨	(١٤-٣) معامل الارتباط الجزئي من الرتبة الأولى
٣١٠	(١٥-٣) معامل الارتباط الجزئي من الرتبة الثانية

#### الفصل الرابع : المخالفات في فروض نموذج الإحدار الخطي المتعدد كيفية اكتشافها وتصحيحها

٣١٤	(١٠٤) مقدمة
٣١٤	(٢٠٤) رسوم البوالي
٣١٥	(٢٠٤) رسوم البوالي الجزئية
٣٢١	(٤٠٤) رسوم الإحدار الجزئي
٣٢٢	(٥٠٤) البوالي المعيارية وبوالي سيكويوننت
٣٢٥	(٦٠٤) استخدام مصفوفة التبع $H$ للتعرف على مشاهدات قاصيه خاصة بالمتغيرات المستقلة
٣٢٧	(٧٠٤) استخدام بوالي سيكويوننت المطبوعة للتعرف على مشاهدات قاصيه خاصة بالمتغير التابع
٣٣٠	(٨٠٤) تحديد المشاهدات المؤثرة
٣٣٠	(١٠٨٠٤) التأثير على القيم المقدرة
٣٣٢	(٢٠٨٠٤) تأثير على معاملات الإحدار
٣٣٧	(٩٠٤) مشكلة عدم الخطية ومعالجتها
٣٤٠	(١٠٠٤) مشكلة عدم تجانس الخطأ ومعالجتها

#### الفصل الخامس : اختبار الفرض نموذج الحدار

٣٥٠	(١٠٥) مقدمة
٣٥١	(٢٠٥) معدل التحديد للمحل
٣٥٢	(٣٠٥) إحصاء ملاوس $C_p$
٣٥٣	(٤٠٥) متوسط مجموع مربعات البوالي
٣٥٤	(٥٠٥) المقاييس $PRESS_p$
٣٥٥	(٦٠٥) طريقة كل الإحدارات الممكنة
٣٦٥	(٧٠٥) طريقة الحذف الخلفي (العكسي)
٣٦٧	(٨٠٥) طريقة الاختيار الأملي (المباشر)
٣٧١	(٩٠٥) طريقة الاختيار لتكرهجي

#### الفصل السادس : نماذج إحدار كثيرات الحدود

٣٧٤	(١٠٦) نماذج إحدار كثيرات الحدود - متغير مستقل واحد
-----	--

٣٧٨	(١-١-٦) تقدير المعالم باستخدام المربعات الصغرى
٣٨١	(٢-١-٦) اختبارات الفروض
٣٩٠	(٣-١-٦) تحديد درجة لنموذج
٣٩٩	(٤-١-٦) تحديد للقيم المتلى
٤٠٠	(٥-١-٦) اختبار بدلالة الاحتمالات
٤٠٣	(٦-١-٦) كثافات الحدود المتعامدة
٤٢٣	(٧-٦) نماذج لحدود كثيرات الحدود – متغيرين مستقلين

## الفصل السابع : المتغيرات الصورية

٤٣٥	(١-٧) المتغيرات الصورية في حالة متغيرات مستقلة وصفية
٤٣٨	(٢-٧) متغير مستقل وصفي بمستويين
٤٤٩	(٣-٧) متغير مستقل وصفي بأكثر من مستويين
٤٥٤	(٤-٧) حالة أكثر من متغير صوري في نموذج الاحددر
٤٥٦	(٥-٧) تطبيقات المتغيرات الصورية في السلاسل الزمنية
٤٥٩	(٦-٧) نماذج الاحددر بمتغيرات صورية تخص متغير الاستجابة
٤٦٠	(١-٦-٧) النموذج الخطي
٤٦٥	(٢-٦-٧) النموذج الغير خطي

## الفصل الثامن : الارتباط الذاتي

٤٧٢	(١-٨) مقدمة
٤٧٣	(٢-٨) أسباب الارتباط الذاتي
٤٧٣	(٣-٨) اختبار درين _ ولسون
٤٧٩	(٤ - ٨) معالجة الارتباط الذاتي
٤٧٩	(١-٤-٨) الطريقة الأولى
٤٨٣	(٢-٤-٨) الطريقة الثانية
٤٨٥	(٣-٤-٨) الطريقة الثالثة
٤٩٦	(٤-٤-٨) الطريقة الرابعة

## الفصل التاسع : الارتباط الخطي المتعدد

٥٠٥	(١-٩) مقدمة
٥٠٦	(٢-٩) مصادر الارتباط الخطي المتعدد
٥٠٧	(٣-٩) تأثيرات الارتباط الخطي المتعدد
٥٠٩	(٤-٩) مؤشرات لوجود الارتباط الخطي المتعدد



٥١٠	(٥.٩) طرق الكشف عن الارتباط الخطي لمتعدد
٥١١	(١.٥-٩) فحص مصفوفة الارتباط
٥١٢	(٢.٥-٩) عوامل تضخم التباين
٥١٥	(٣.٥-٩) تحليل القيم المميزة
٥١٧	(٤.٥-٩) تشخيصات أخرى
٥٢٣	(٦.٩) معالجة الارتباط الخطي لمتعدد
٥٣٠	(٧.٩) تحديد المكونات الرئيسية

## الفصل العاشر: نماذج الانحدار الغير خطية

٥٣٨	(١.١٠) مقدمة
٥٣٨	(٢.١٠) نموذج الانحدار الغير خطي
٥٣٩	(٣.١٠) لمربعات الصغرى الغير خطية
٥٤١	(٤.١٠) لتحويل إلى نموذج خطي
٥٤٥	(٥.١٠) تقدير المعامل في نظام غير خطي
٥٥٣	(٦.١٠) لقيم المبذوبة
٥٥٧	(٧.١٠) أمثلة للنماذج الغير خطية
٥٥٨	المراجع
٥٦٤	الملحق



# الفصل الأول

## الانحدار الخطي البسيط

### Simple Linear Regression

مفاهيم أساسية	(١-١)
العلاقة بين متغير مستقل ومتغير تابع	(٢-١)
العلاقة للدالية	(١-٢-١)
العلاقة الإحصائية	(٢-٢-١)
مقدمة في الانحدار الخطي البسيط	(٣-١)
نموذج الانحدار الخطي البسيط	(٤-١)
فروض نموذج الانحدار الخطي البسيط	(٥-١)
طريقة المربعات الصغرى	(٦-١)
خواص مقدرات المربعات الصغرى	(٧-١)
صيغة بديله للنموذج	(٨-١)
تقدير $\sigma^2$	(٩-١)
استدلالات تخص معاملات الانحدار	(١٠-١)
التنبؤ	(١١-١)
أسلوب تحليل الانحدار	(١٢-١)
معامل التحديد	(١٣-١)
الانحدار خلال نقطة الأصل	(١٤-١)
الاستدلال آنياً لمعالم النموذج	(١٥-١)
التقدير آنياً لمتوسط الاستجابة	(١٦-١)
التنبؤ لعدد $m$ من مشاهدات جديدة	(١٧-١)
التقدير باستخدام الإمكان الأعظم	(١٨-١)
الارتباط	(١٩-١)

## (١-١) مفاهيم أساسية

يهتم تحليل الانحدار بالعلاقة بين متغير موضوع الدراسة، يسمى المتغير التابع أو متغير استجابة response variable وواحد أو أكثر من متغيرات أخرى تسمى متغيرات مستقلة independent variables أو متغيرات مفسره explanatory variables أو متغيرات تتنبؤ predictor variables .

### أمثلة

١. اختار باحث تغذية أربعة نساء عشوائيا من كل شريحة عمرية من 10 سنوات تبدأ بالعمر 40 وتنتهي بالعمر 79 وكان المتغير التابع  $Y$  هو قياس كتلة العضلة، أما المتغير المستقل فكان العمر ( $x$ ).

٢. في دراسة أجريت في مؤسسه علميه كان المتغير التابع ( $Y$ ) هو الرواتب السنوية لمباحثين في الرياضيات من مستوى متوسط ومتقدم ( $Y$  بالآلاف للدولارات) أما المتغيرات المستقلة فكانت رقم قياسي يعبر عن النجاح في الحصول على دعم منه ( $x_1$ ) وعدد سنوات الخبرة ( $x_2$ ).

غالبا ما يستخدم تحليل الانحدار في التنبؤ بالمتغير التابع من المعلومات عن واحد أو أكثر من المتغيرات المستقلة. في هذا الفصل والفصل القادم سوف نقدم بعض المفاهيم الأساسية وطرق الاستدلال لتحليل الانحدار البسيط حيث يعتمد المتغير التابع على متغير مستقل واحد. وفي الفصل الثالث سوف نتناول الانحدار المتعدد حيث يعتمد المتغير التابع على متغيرين مستقلين أو أكثر. وتجدر الإشارة هنا إن كلمة الانحدار استخدمت لأول مرة بصيغتها الحاضرة من قبل عالم الوراثة البريطاني السير فرانسيس كالتون (Sir Francis Galton) حيث كان واحد من أول الباحثين الذين تعاملوا مع موضوع دراسة أو وصف متغير واحد بالاعتماد على ولعد أو أكثر من المتغيرات. فقد درس كالتون العلاقة بين أطوال الأبناء مقارنة بأطوال آبائهم فلاحظ وجود علاقة واضحة وهي ميل أطوال الأبناء نحو المتوسط لأطوال آبائهم. فالأبناء قصار القامة يميلون لإنتاج أبناء متوسط أطوالهم أعلى (أطول من آبائهم). بينما العكس صحيح في حالة الإبناء طوال القامة بشكل غير اعتيادي. لذلك فإن العالم كالتون ذكر أن أطوال أبناء آباء طوال أو قصار تبدو وكأنها "ترتد" أو تتحدر (regress) نحو المتوسط للمجموعة ولذلك ظهرت كلمة الانحدار regression. وقد نشرت هذه النتيجة الدراسية في عام 1885 تحت عنوان "regression toward mediocrity in hereditary stature". من تلك البداية فإن كلمة الانحدار قد طورت إلى المعنى الذي يشمل تحليل البيانات التي تحتوي على اثنين أو أكثر من المتغيرات عندما يكون الهدف هو اكتشاف طبيعة هذه العلاقة وبعد ذلك اعتمادها في قضايا البحث العلمي.

### (٢-١) العلاقة بين متغير مستقل ومتغير تابع

#### Relation between dependent and independent variable

عند دراسة العلاقة بين متغير تابع  $Y$  ومتغير مستقل  $x$  يكون من المفيد التمييز بين العلاقة الدالية والعلاقة الإحصائية.

### (١-٢-١) العلاقة الدالية

#### Functional relation

العلاقات التي يكون فيها تقدير  $Y$  وحيد وذلك من المعلومات عن  $x$  تسمى العلاقات الدالية.

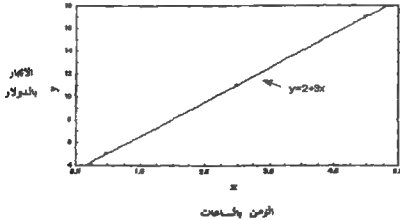
تعريف: العلاقة الدالية بين متغير تابع  $Y$  ومتغير مستقل  $x$  تمثل علاقة مضبوطة exact حيث قيمة  $Y$  التي تقدر تكون وحيدة عندما تحدد قيمة للمتغير  $x$ .

#### أمثلة

١. الإيجار ( $Y$  بالدولار) لموتور كهربائي يرتبط بعدد ساعات تاجيره ( $x$ ) كما يأتي:

$$Y = 2 + 3x$$

حيث 2 قيمة ثابتة على الفتورة و 3 تمثل مبلغاً مضافاً لكل ساعة إيجار. وعلى ذلك لأي عدد من الساعات هناك قيمة وحيدة للإيجار. يوضح شكل (١-١) خط العلاقة  $Y = 2 + 3x$  وأيضاً المشاهدات لثلاثة مبالغ مدفوعة لتأجير 1, 3, 5 ساعات على التوالي. النقاط على الرسم سوف تكون (1,5), (3,11), (5,17).



شكل (١-١)

وعلى ذلك قيمة  $Y$  التي تقدر من  $x$  تكون وحيدة وكل المشاهدات تقع على خط العلاقة.

٢- إذا كانت السرعة الأولية لجزئ هي  $v_0$  وإذا كانت  $a$  هو ثابت التعجيل (الإسراع) فإن المسافة المقطوعة ( $Y$ ) تحسب من المعادلة التالية:

$$Y = v_0 x + \frac{1}{2} ax^2$$

حيث  $x$  تمثل الزمن .

٣- إذا كانت قيمة التكلفة بالطائرة تحدد على أساس قيمة ثابتة مقدارها 50 مضافا له مبلغا مقداره 0.10 دولار لكل كيلو متر من المسافة المقطوعة وإذا كانت  $Y$  هي قيمة التكلفة بالطائرة و  $x$  هي عدد الكيلومترات المقطوعة فإن العلاقة بين  $Y$  ,  $x$  تحدها المعادلة التالية:

$$Y = 50 + 0.1x$$

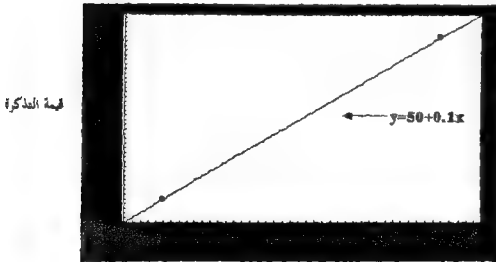
فعندما تكون مسافة الرحلة هي 300 كم فإن قيمة التكلفة بالطائرة هي :

$$Y = 50 + 0.1(300) = 80$$

وعندما تكون مسافة الرحلة 1000 كم فإن قيمة التكلفة بالطائرة هي :

$$Y = 50 + 0.1(1000) = 150$$

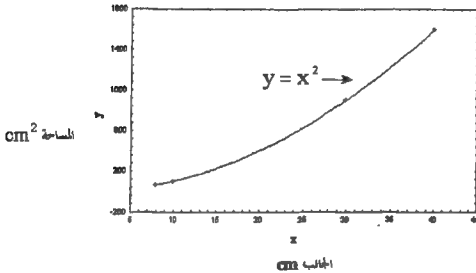
يوضح شكل (٢-١) خط العلاقة  $Y = 50 + 0.1x$ .



مسافة الرحلة

شكل (٢-١)

٤- مساحة الألواح المربعة من معدن  $(Y \text{ cm}^2)$  ترتبط بطول جانبيه  $(x \text{ سم})$  بعلاقة دالية  $Y = x^2$ . يوضح شكل (١-٣) منحنى العلاقة وأيضا مشاهدات لأربعة ألواح جوانبها هي 8, 10, 30, 40 سم على التوالي.



شكل (١-٣)

### (١-٢-٢) العلاقة الإحصائية

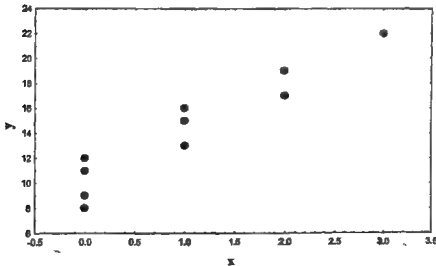
#### Statistical relation

في كثير من الدراسات التطبيقية، فإن قيمة المتغير التابع  $Y$  لا تقدر بقيمة وحيدة وذلك عندما تحدد قيمة خاصة للمتغير المستقل  $x$ . فعلى سبيل المثال، عند دراسة العلاقة بين دخل الأسرة وإنفاقها على الطعام، فإننا نجد أسر لها نفس مستوى الدخل تختلف في إنفاقها على الطعام، السبب الرئيسي لذلك هو وجود عوامل أخرى غير دخل الأسرة تلعب دوراً، مثل حجم الأسرة ونظام المعيشة...الخ.

العلاقات التي يكون فيها تقدير المتغير التابع  $Y$  ليس وحيداً وذلك من المعلومات عن المتغير المستقل  $x$  تسمى علاقات إحصائية.

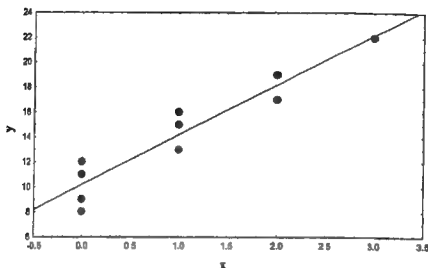
تعريف: العلاقات الإحصائية بين متغير تابع  $Y$  ومتغير مستقل  $x$  تمثل علاقة غير مضبوطة  $inexact \text{ relation}$  حيث تقدير قيمة  $Y$  لا تكون وحيدة عندما تحدد قيمة لـ  $x$ .

بفرض متغيرين بينهما علاقة إحصائية فإنه لقيمة ثابتة من  $x$  فإن القيمة للمتغير  $Y$  سوف تكون عشوائية random، أي أن  $Y$  متغير عشوائي. فعلى سبيل المثال إذا كان الاهتمام بدراسة العلاقة بين عمر الطفل ( $x$ ) وحجم المفردات التي يتعلمها ( $Y$ ) وبفرض أننا أجرينا اختبار لطفل عمره خمس سنوات  $x=5$  فإن حجم المفردات التي يتعلمها الطفل قبل إجراء الاختبار تمثل متغير عشوائي. ولكن بعد اختبار طفل عمره خمس سنوات وتسجيل عدد الجمل التي تعلمها فقد تكون مثلاً 2000. في هذه الحالة نقول أن القيمة المشاهدة لـ  $Y$  والمرتبطة بقيمه  $x=5$  هي  $Y=2000$ . كمثال آخر لعلاقة إحصائية، وبفرض أن مادة ما تستخدم في الأبحاث الحيوية والطبية تشحن إلى المستخدمين جواً وذلك في صناديق تحتوي كل منها 1000 أنبوبة، والبيانات التي تم جمعها تناولت عشر شحنات، حيث  $x$  تمثل عدد المرات التي يحول فيها الصندوق من طائرة إلى أخرى خلال خط سير الشحنة، و  $Y$  تمثل عدد الأنبوبات التي وجدت مكسورة عند وصولها. ففي الشحنة الأولى كان  $x=1$  و  $y=16$  وعلى ذلك النقطة لهذه الشحنة توقع على الرسم عند (1,16) كما هو موضح في شكل (٤-١). النقاط الأخرى توقع على الرسم بنفس الشكل. يتضح من الرسم أن عدد الأنبوبات التي وجدت مكسورة عند وصولها يزيد كلما زادت عدد المرات التي يحول فيها الصندوق من طائرة إلى أخرى. لوصف هذا الاتجاه فإننا نقيم خط مستقيم يمر خلال النقاط كما هو موضح في شكل (٥-١). إذا العلاقة إحصائية وذلك لأن قيمة  $Y$  تختلف عن نفس القيمة من  $x$ . العلاقة الإحصائية هنا خطية أي تتبع خط مستقيم.



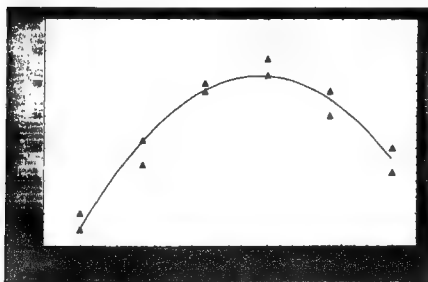
شكل (٤-١)





شكل (٥-١)

الآن في تجربة لدراسة فاعلية (جير) تجريبي جديد في تخفيض استهلاك الجازولين في 12 محاولة استخدمت فيها عربة نقل خفيفة مجهزة بهذا الجير. في هذه التجربة كان المتغير المستقل  $x$  هو السرعة الثابتة (بالميل في الساعة) لعربة الاختبار والمتغير التابع  $Y$  هنا هو عدد الأميال المقطوعة لكل جالون. شكل الانتشار لبيانات هذه التجربة موضح في شكل (٦-١) حيث العلاقة هنا إحصائية وعلى شكل منحنى.



شكل (٦-١)

تتضح من المثالين السابقين خاصيتين للعلاقة الإحصائية:

١. يتجه المتغير التابع  $Y$  للتغير بنظام معين مع المتغير المستقل  $x$  والذي يوصف بالخطي أو بالمنحني.
٢. انتشار المشاهدات حول الخط أو المنحني للعلاقة الإحصائية يرجع جزئياً إلى عوامل أخرى غير تأثير المتغير المستقل  $x$  على المتغير التابع  $Y$ .

في هذا الفصل سوف تقتصر دراستنا على الانحدار الخطي البسيط والذي يعني أن المتغير التابع يعتمد على متغير مستقل واحد وأن العلاقة بينهما يمكن أن تمثل بمعادلة خط مستقيم.

### (٣-١) مقدمة في الانحدار الخطي البسيط

يفرض عينة عشوائية من الحجم  $n$  ممثلة بأزواج المشاهدات  $\{(x_i, y_i); i = 1, 2, \dots, n\}$ . عينات متكررة فإننا سوف نأخذ بالضبط قيم  $x$  ونتوقع تغير في قيم  $y$ . وعلى ذلك قيمة  $y_i$  في الزوج المرتب  $(x_i, y_i)$  تمثل قيمة لمتغير عشوائي  $Y_i$ . أي أن النتيجة التي يأخذها  $Y_i$  غير مؤكدة uncertain ولا يمكن السيطرة عليها بواسطة الباحث. سوف نعرف  $Y|x$  لتمثل متغير عشوائي  $Y$  يقابل قيمة ثابتة  $x$ ، ونعرف متوسطة بالرمز  $\mu_{Y|x}$  وتباينه بالرمز  $\sigma_{Y|x}^2$ . من الواضح أنه عندما  $x = x_i$  فإن الرمز  $Y|x_i$  يمثل المتغير العشوائي  $Y_i$  بمتوسط  $\mu_{Y|x_i}$  وتباين  $\sigma_{Y|x_i}^2$ .

أن الانحدار الخطي البسيط يعني أن  $\mu_{Y|x}$  ترتبط خطياً بـ  $x$  بمعادلة انحدار المجتمع التالية:

$$\mu_{Y|x} = \beta_0 + \beta_1 x$$

حيث معاملات الانحدار  $\beta_0, \beta_1$ ، يمثلان معلمتين مطلوب تقديرهما من مشاهدات العينة حيث  $b_0$  تقدير للمعلمة  $\beta_0$  و  $b_1$  تقدير للمعلمة  $\beta_1$ . أي أننا نقدر  $\mu_{Y|x}$  بـ  $\hat{y}$  من انحدار العينة أو خط الانحدار المقدّر التالي:

$$\hat{y} = b_0 + b_1 x$$

### شكل الانتشار

الأسلوب المفيد ليدعم تحليل الانحدار هو تمثيل البيانات بيانياً وهو ما يعرف بشكل الانتشار scatter plot وذلك من فئة المشاهدات  $\{(x_i, y_i), i=1, 2, \dots, n\}$ . للحصول على شكل الانتشار يخصص محور  $x$  (المحور الأفقي) للمتغير المستقل بينما يخصص محور  $y$  (المحور الرأسي) للمتغير التابع. لكل زوج  $(x, y)$  من أزواج المشاهدات التي عددها  $n$  نقوم بتوقيع نقطة على الرسم. تتوفر كثير من برامج الحاسب الآلي الجاهزة والخاصة بالانحدار مثل برنامج SPSS و Statistica و Minitab للحصول على أشكال الانتشار. يفيد شكل الانتشار فيما يلي :

- يوضح عموماً فيما إذا كانت هناك علاقة ظاهرة بين المتغيرين أم لا .
- عند وجود علاقة يوضح شكل الانتشار فيما إذا كانت العلاقة خطية أم لا .
- إذا كانت العلاقة خطية فإن شكل الانتشار يوضح فيما إذا كانت سلبية (عكسية) أو موجبة (طردية).

### مثال (١-١)

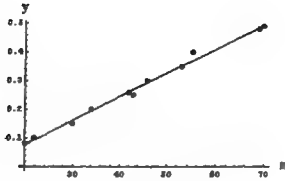
في إحدى التجارب وزن قرون عدد من الغزلان المختلفة الأعمار وكانت النتائج كما هي معطاة في جدول (١-١). المطلوب رسم شكل الانتشار وتحديد شكل العلاقة بين المتغيرين .

جدول (١-١)

المر	20	22	30	34	42	43	46	53	55	69	70
x											
الوزن	0.08	0.10	0.15	0.20	0.26	0.25	0.30	0.35	0.40	0.48	0.49
y											

### الحل

يوضح من شكل (١-١) أن النقاط عموماً ، ليس بالضبط ، تقع على خط مستقيم. هذا يجعلنا نقترح أن العلاقة بين المتغيرين يمكن وصفها (كتقريب أول) بمعادلة خط مستقيم .



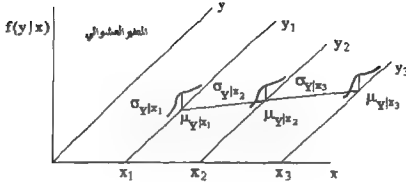
شكل (٧-١)

#### (٤-١) نموذج الانحدار الخطي البسيط

في حالة الانحدار الخطي البسيط حيث يوجد متغير مستقل واحد  $X$  ومتغير تابع  $Y$  فإن البيانات تمثل بأزواج المشاهدات  $\{(x_i, y_i), i = 1, 2, \dots, n\}$ . سنعرف كل متغير عشوائي  $Y_i = Y | x_i$  بنموذج إحصائي Statistical model وذلك تحت فرض أن كل المتوسطات  $\mu_{Y|x_i}$  تقع على خط مستقيم كما هو موضح في شكل (٨-١). وعلى ذلك فإن كل متغير  $Y_i$  يمكن وصفه بنموذج انحدار بسيط كالتالي:

$$Y_i = \mu_{Y|x_i} + \varepsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (١-١)$$

حيث المتغير العشوائي  $\varepsilon_i$  ، خطأ النموذج ، لابد أن يكون له متوسط يساوي صفر.



شكل (٨-١)

تشير المعلمة  $\beta_1$  في نموذج الانحدار (١-١) (والتي هي ميل خط الانحدار) إلى التغير في متوسط التوزيع الاحتمالي للمتغير التابع  $Y$  لكل وحدة زيادة في  $x$ . أما المعلمة  $\beta_0$  فتمثل للتقاطع الصادي لخط الانحدار. وإذا احتوى مدى النموذج على القيمة  $x = 0$  فإن  $\beta_0$  تعطي متوسط التوزيع الاحتمالي لمتغير  $Y$  عندما  $x = 0$ . وليس للمعلمة  $\beta_0$  أي تفسير خاص بها كحد منفصل في نموذج الانحدار إذا لم يتضمن مجاله القيمة  $x = 0$ .

يقال عن النموذج (١-١) انه بسيط وخطي في المعالم وخطي في المتغير المستقل. فهو بسيط لأنه يستخدم متغيراً مستقلاً واحداً فقط، وخطي في المعالم لأنه لا تظهر أي معلمة كاس أو مضروبة بمعلمة أخرى، وخطي في المتغير المستقل لأن هذا المتغير لا يظهر إلا مرفوعاً للأس الواحد. أيضاً يعرف النموذج (١-١) بالنموذج من الرتبة الأولى والذي يختلف عن النموذج البسيط التالي:

$$Y_i = \beta_0 + \beta_1 x_i^2 + \varepsilon_i$$

والذي يكون خطي في المعالم وغير خطي في المتغير المستقل لأن هذا المتغير يظهر مرفوعاً للأس 2 ويمثل نموذج خطي في المعالم ومن الرتبة الثانية في  $x$ .

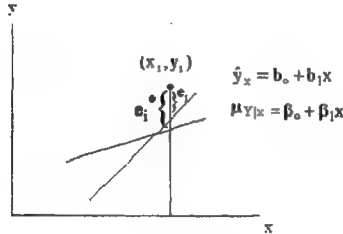
كل مشاهدة  $(x_i, y_i)$  في عينة عشوائية من الحجم  $n$  تحقق العلاقة :

$$y_i = \beta_0 + \beta_1 x_i + e_i^*$$

حيث  $e_i^*$  قيمة مفترضة للمتغير  $e_i$  عندما تأخذ القيمة  $y_i$ . المعادلة السابقة ينظر إليها كنموذج لمشاهدته مفردة  $y_i$ . بنفس الشكل ، باستخدام معادلة خط الانحدار المقدرة فإن :

$$y_i = b_0 + b_1 x_i + e_i,$$

حيث  $e_i = y_i - \hat{y}_i$  تسمى البالي residual والذي يصف خطأ في توفيق النموذج عند نقطة الملاحظة رقم  $i$ . الفرق بين  $e_i$  و  $e_i^*$  و موضع في شكل (٩-١)، يوضح شكل (٩-١) الخط المقدر من فئة البيانات والمسمى  $\hat{y} = b_0 + b_1 x$  وخط الانحدار الحقيقي  $\mu_{Y|X} = \beta_0 + \beta_1 x$ . الآن بالطبع  $\beta_0, \beta_1$  معلمتين غير معلومتين. يعتبر الخط المقدر تقدير للخط  $\mu_{Y|X}$ . ومما يجدر الإشارة إليه أن  $e_i$  يمكن ملاحظتها، أما  $e_i^*$  فلا يمكن ملاحظتها لأن الخط  $\mu_{Y|X}$  مفترض وغير معروف.



شكل (٩-١)

#### (٥-١) فروض نموذج الانحدار الخطي البسيط

لتقدير معالم نموذج الانحدار (١-١) توضع الفروض التالية لحد الخطأ  $e_i$  والمسماة فروض جلوس - ماركوف Gauss-Markov.

$$E(e_i) = 0,$$

$$E(e_i^2) = \sigma^2, E(e_i e_j) = 0$$

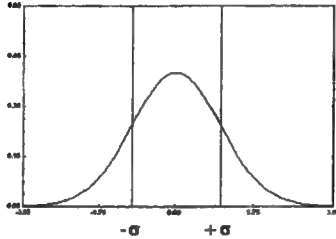
حيث  $j \neq i$  لكل  $i, j = 1, \dots, n$  أي أن  $\varepsilon_i, \varepsilon_j$  غير مرتبطين.  
وعلى ذلك:

$$E(Y_i) = \beta_0 + \beta_1 x_i, \text{Var}(Y_i) = \sigma^2.$$

هناك فروض أخرى نحتاج لها عند إجراء فترات ثقة واختبارات فروض تخص المعلمتين  $\beta_0, \beta_1$  وهي أن  $\varepsilon_i$  يتبع التوزيع الطبيعي بمتوسط صفر وتباين  $\sigma^2$ ، أي أن:

$$\varepsilon_i \sim N(0, \sigma^2).$$

توزيع  $\varepsilon_i$  موضح في شكل (١-١) .



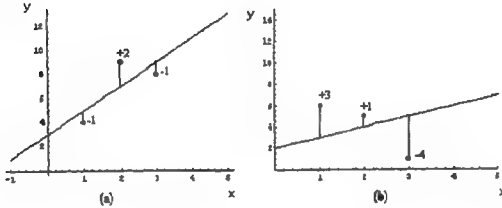
شكل (١-١)

#### (١-٦) طريقة المربعات الصغرى

##### The method of least squares

بالرغم من وجود العديد من الطرق للحصول على تقديرات للمعلمتين  $\beta_0, \beta_1$  إلا أن أفضل هذه الطرق هي طريقة المربعات الصغرى. ترجع هذه الطريقة إلى عالم الرياضيات الألماني كارل فريدريكس جوس Carl Friedrich Gauss . وبما أن الخط المطلوب يكون لأغراض التنبؤ لذلك من المناسب أن يكون الخط من الدقة بحيث تكون أخطاء التقدير صغيرة. والمقصود هنا بأخطاء التقدير الفروق بين القيم المشاهدة  $y_i$  والقيم المناظرة  $\hat{y}_i$  (البنوافي) على الخط المستقيم. أي أن أخطاء التقدير هي  $(y_i - \hat{y}_i)$ ,  $i = 1, 2, \dots, n$ . أخطاء التقدير موضحة في شكل

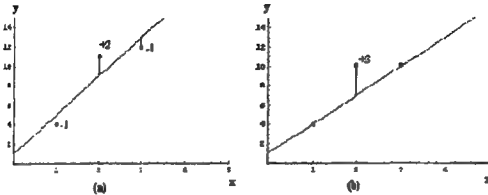
(١١-١) a وشكل (١١-١) b بأجزاء الخطوط الرأسية التي تحصل بين النقاط والخط المستقيم. النقطة الواقعة فوق الخط تعطي خطأ (باقي) موجب والنقطة الواقعة تحت الخط تعطي خطأ سالب. ولحد من الطرق لتقليل الأخطاء هو جعل  $\sum_{i=1}^n (y_i - \hat{y}_i)$  أقل ما يمكن ، ولكن جعل  $\sum_{i=1}^n (y_i - \hat{y}_i)$  أقل ما يمكن لا يعني الحصول على توفيق جيد. ففي شكل (١١-١) a ثلاثة أخطاء واحد موجب والآخرين سالبين حيث  $\sum_{i=1}^n (y_i - \hat{y}_i) = 0$  في هذه الحالة بتقليل الخطأ فإننا حصلنا على توفيق يبدو جيد.



شكل (١١-١)

الآن بالنظر إلى شكل (١١-١) b فإن خط الانحدار أدى إلى جعل  $\sum_{i=1}^n (y_i - \hat{y}_i) = 0$  وبالرغم من ذلك يتضح أن التوفيق رديء. الآن ماذا يحدث عند إهمال الإشارة وإيجاد الخط المقدر الذي يجعل  $\sum_{i=1}^n |y_i - \hat{y}_i|$  أقل ما يمكن ؟ مرة أخرى لم نضمن أن الخط يمثل أفضل توفيق. في شكل (١٢-١) يتضح أن الخط في (a) أفضل من الخط في (b) بالرغم من أن الخط في (b) جعل  $\sum_{i=1}^n |y_i - \hat{y}_i|$  أقل من (a)





شكل (١٢-١)

وعلى ذلك نجد أن استخدام القيم المطلقة ليس مناسباً في المعالجة الرياضية وذلك فإن هذه الصعوبة يمكن تلافيها بأن نطلب أن يكون مجموع مربعات الأخطاء صغيراً بقدر الإمكان. قيم المعالم هذه التي تقلل إلى أقصى حد مجموع مربعات الأخطاء تحدد ما يعرف بالفضل خط مستقيم يوفق النقاط المشاهدة من جهة نظر المربعات الصغرى. ومما يجدر الإشارة إليه أن طريقة المربعات الصغرى لتوفيق خط مستقيم لمجموعه من النقاط يمكن تطبيقها سواء كانت قيم  $x$  حددت مسبقاً أو تمثل قيم لمتغير عشوائي، أي إذا كان المتغير المستقل والمتغير التابع يمثلان متغيرات عشوائية. وفي هذه الحالة تطبق طريقة المربعات الصغرى إذا تحقق الشرطان التاليان :-

١. للتوزيعات الشرطية للمتغيرات التابعة  $Y_i$  علماً بأن  $x_i$  معطاة تمثل توزيعات طبيعية مستقلة لها متوسط شرطي  $\beta_0 + \beta_1 x_i$  وتباين شرطي  $\sigma^2$ .
  ٢. المتغيرات  $X_i$  هي متغيرات عشوائية مستقلة وتوزيعها الاحتمالي  $g(x_i)$  لا يحتوي على المعالم  $\beta_0, \beta_1, \sigma^2$ .
- الآن سوف نوضح مفهوم المربعات الصغرى بالمثال التالي :

مثال (٢-١)

أجريت تجربة لدراسة العلاقة بين التسميد ومحصول الذرة، والبيانات التي تم الحصول عليها معطاة في جدول (٢-١).

جدول (٢-١)

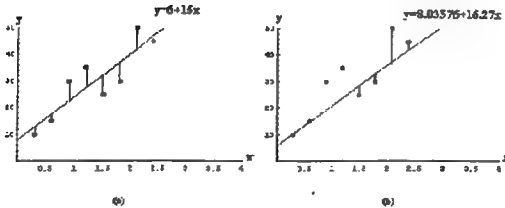
x	y	x <sup>2</sup>	xy
0.3	10	0.09	3
0.6	15	0.36	9
0.9	30	0.81	27
1.2	35	1.44	42
1.5	25	2.25	37.5
1.8	30	3.24	54
2.1	50	4.41	105
2.4	45	5.76	108
10.8	240	18.36	385.5

شكل الانتشار للبيانات المعطاة في جدول (٢-١) موضح في شكل (١٣-١). يتضح في شكل (١٣-١) أن النقاط عموماً "ليست بالضبط" تقترب من خط مستقيم وهذا يجعلنا نقترح أن العلاقة بين المتغيرين يمكن وصفها "كتقريب أولي" بمعادلة خط مستقيم. الخط المقدر  $y = 6 + 15x$  موضح على نفس الرسم. المسافات الرأسية أو الانحرافات "deviations" للملاحظات عن الخط المقدر موضحة في شكل (١٣-١) a. على سبيل المثال للحالة الأولى حيث  $x_1 = 0.3, y_1 = 10$  فإن القيمة المقدرة هي :

$$\hat{y}_1 = 6 + 15(0.3) = 10.5$$

والانحراف الراسي هو :

$$y_1 - \hat{y}_1 = -0.5$$



شكل (١٣-١)

زيادة الانحرافات الرأسية عن الخط المقدر تعتبر مؤشرا لرداءة التوفيق. إن طريقة المربعات الصغرى تقيس جودة التوفيق للخط المستقيم وذلك من مجموع مربعات الانحرافات  $\sum (y_i - \hat{y}_i)^2$ . سوف نرمز لمجموع مربعات الانحرافات بالرمز SSE. للخط الموفق في شكل (١٣-١) فإن قيمة SSE هي :

$$SSE = (10 - 10.5)^2 + (15 - 15)^2 + \dots + (45 - 42)^2 = 418 .$$

أفضل خط مقدر هو الموضح في شكل (١٣-١) وهو :

$$\hat{y} = 8.03571 + 16.2698x$$

لهذا الخط المقدر فإن ناتج مجموع مربعات الانحرافات هو :

$$SSE = (10 - 12.9167)^2 + (15 - 17.7976)^2 + \dots + (45 - 47.0833)^2 = 299.405 .$$

في الحقيقة فإن الخط في شكل (١٣-١) هو الخط الذي له أقل مجموع مربعات للانحرافات بين كل الخطوط المقدرة من فئة المشاهدات. إن طريقة المربعات الصغرى تعتبر الطريقة التي تعطي أفضل خط مقدر بحيث أن SSE أقل ما يمكن. وعلى ذلك

$$\hat{y} = 8.03571 + 16.2698x .$$

هو خط الانحدار لطريقة المربعات الصغرى حيث (8.03571) هو تقدير للمعلمة  $\beta_0$  و (16.2698) هو تقدير للمعلمة  $\beta_1$  .

تتطلب طريقة المربعات الصغرى الحصول على التقديرين  $b_0, b_1$  وذلك للمعلمتين  $\beta_0, \beta_1$  على التوالي اللذين يجعلان مجموع مربعات الأخطاء (البواقي) SSE أقل ما يمكن، أي اللذين يحققان النهاية الصغرى لمجموع مربعات البواقي، حيث يعرف مجموع مربعات البواقي كالآتي :

$$SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 .$$

بإجراء التفاضل الجزئي لـ SSE بالنسبة لكل من  $b_0, b_1$  نحصل على :

$$\frac{\partial(SSE)}{\partial b_0} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) \quad (٢-١)$$

$$\frac{\partial(SSE)}{\partial b_1} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) x_i . \quad (٣-١)$$

بمساواة المعادلة (٢-١) بالصفر نحصل على :

$$\sum_{i=1}^n (y_i - b_0 - b_1 x_i) = 0.$$

أي :

$$\sum_{i=1}^n y_i = nb_0 + b_1 \sum_{i=1}^n x_i. \quad (٤-١)$$

وبمساواة المعادلة (٣-١) بالصفر وإعادة تنظيم المعادلة نحصل على :

$$\sum_{i=1}^n x_i y_i = b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2. \quad (٥-١)$$

يتم الحصول على التقديرين  $b_0, b_1$  بحل المعادلتين الطيبميتين (٤-١) و (٥-١) آنيا.

ويمكن توضيح ذلك في المثال الآتي :

مثال (٣-١)

نفرض أنه تم دراسة العلاقة بين مصاريف الإعلان لسلعة ما  $x$  (£000) والمبيعات للسلعة  $y$  (£m) والبيانات موضحة في جدول (٣-١).

جدول (٣-١)

x	y	$x^2$	xy
100	9	10000	900
105	8	11025	840
90	5	8100	450
80	2	6400	160
80	4	6400	320
85	6	7225	510
87	4	7569	348
92	7	8464	644
90	6	8100	540
95	7	9025	665
93	5	8649	465
85	5	7225	425
85	4	7225	340
70	3	4900	210
85	3	7225	255
1322	78	117532	7072

ومن جدول (٣-١) يمكن حساب قيمة  $b_0, b_1$  كالآتي :

$$78 = 15b_0 + 1322b_1$$

$$7072 = 1322b_0 + 117532b_1.$$

بضرب المعادلة الأولى في 1322 (معامل  $b_0$  في المعادلة الثانية) نحصل على:

$$103116 = 19830b_0 + 1747684b_1.$$

الآن بضرب المعادلة الثانية في 15 (معامل  $b_0$  في المعادلة الأولى) نحصل على :

$$106080 = 19830b_0 + 1762980b_1.$$

الآن سوف يكون لدينا زوج من المعادلات الآتية حيث معامل  $b_0$  واحد في الاثنين. بطرح المعادلتين نحصل على معادلة لا تحتوي على الحد  $b_0$  :

$$106080 = 19830b_0 + 1762980b_1$$

$$103116 = 19830b_0 + 1747684b_1$$

$$2964 = 15296b_1$$

$$b_1 = 0.193776 .$$

الآن بعد الحصول على قيمة  $b_1$  نعوض عنها في المعادلة الأولى وذلك لإيجاد قيمة  $b_0$  :

$$78 = 15b_0 + (0.193776)(1322)$$

$$78 = 15b_0 + 256.1721$$

$$-178.1721 = 15b_0$$

$$b_0 = -11.8781 .$$

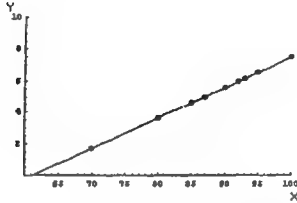
وعلى ذلك معادلة الاتحاد المقدرة سوف تكون :

$$\hat{y} = -11.8781 + 0.19378x$$

أو بصورة بسيطة..

$$\hat{y} = -11.9 + 0.19x.$$

والموضحة بيانياً في شكل (١٤-١) مع شكل الانتشار.



شكل (١٤-١)

يتضح في المثال السابق أن طريقة حساب معادلة الانحدار المقدر بحل المعادلات الطبيعية آتياً عملية صعبة وعلى ذلك يمكن إيجاد الصيغ الحسابية لكل من  $b_0, b_1$  من المعادلات الطبيعية (٤-١)، (٥-١) كالآتي : بضرب المعادلة

(٤-١) في  $\sum_{i=1}^n x_i$ ، وضرب المعادلة (٥-١) في  $n$  نحصل على :

$$\sum_{i=1}^n x_i \sum_{i=1}^n y_i = nb_0 \sum_{i=1}^n x_i + b_1 \left( \sum_{i=1}^n x_i \right)^2, \quad (٦-١)$$

$$n \sum_{i=1}^n x_i y_i = nb_0 \sum_{i=1}^n x_i + b_1 n \sum_{i=1}^n x_i^2. \quad (٧-١)$$

ويطرح المعادلة (٦-١) من المعادلة (٧-١) نحصل على :

$$\begin{aligned} n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i &= b_1 n \sum_{i=1}^n x_i^2 - b_1 \left( \sum_{i=1}^n x_i \right)^2 \\ &= b_1 \left[ n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] \end{aligned}$$

وعليه فإن :

$$b_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left( \sum x_i \right)^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (٨-١)$$

وبقسمة طرفي المعادلة (٤-١) على n نحصل على :

$$b_0 = \bar{y} - b_1 \bar{x} \quad (٩-١)$$

حيث  $\bar{x}, \bar{y}$  يرمزان للوسط الحسابي للعينة للمتغير المستقل  $x$  والمتغير التابع  $Y$  على التوالي. وعلى ذلك التقديرين  $b_0, b_1$  يمكن الحصول عليهما بإيجاد الانحرافات حول المتوسط  $(x_i - \bar{x}), (y_i - \bar{y})$  لملاحظات العينة ثم حساب الكمية  $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$  و  $\sum_{i=1}^n (x_i - \bar{x})^2$  وبالتعويض عليهما في (٨-١) نحصل على  $b_1$  أما  $b_0$  فنحصل عليها من المعادلة (٩-١).

لبيانات المثال (٣-١) والمعطاة في جدول (٣-١) المطلوب حساب قيمة  $b_1$  من المعادلة (٨-١) وقيمة  $b_0$  من المعادلة (٩-١) ثم إيجاد معادلة الانحدار المقردة .

الحل

من جدول (٣-١) نحصل على جدول (٤-١) .

جدول (٤-١)

x	y	(x - $\bar{x}$ )	(y - $\bar{y}$ )	(x - $\bar{x}$ )(y - $\bar{y}$ )	(x - $\bar{x}$ ) <sup>2</sup>
100	9	11.8667	3.8	45.0933	140.818
105	8	16.8667	2.8	47.2267	284.484
90	5	1.86667	-0.2	-0.373333	3.48444
80	2	-8.13333	-3.2	26.0267	66.1511
80	4	-8.13333	-1.2	9.76	66.1511
85	6	-3.13333	0.8	-2.50667	9.81778
87	4	-1.3333	-1.2	1.36	1.28444
92	7	3.86667	1.8	6.96	14.9511
90	6	1.86667	0.8	1.49333	3.48444
95	7	6.86667	1.8	12.36	47.1511
93	5	4.86667	-0.2	-0.973333	23.6844
85	5	-3.13333	-0.2	0.626667	9.81778
85	4	-3.13333	-1.2	3.76	9.81778
70	3	-18.1333	-2.2	39.8933	328.818
85	3	-3.13333	-2.2	6.89333	9.81778
1322	78			197.6	1019.73

من جدول (٤-١) نحصل على :

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1322}{15} = 88.133 \quad , \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{78}{15} = 5.20 .$$

الانحرافات عن متوسط الملاحظة الأولى هو :

$$(x_1 - \bar{x}) = (100 - 88.133) = 11.8667 \quad , \quad (y_1 - \bar{y}) = (9 - 5.20) = 3.8 .$$

الانحرافات للملاحظات الأخرى تحسب بنفس الطريقة المعطاة في العمود الثالث والرابع على اليسار من جدول (٤-١). الآن نحسب الصيغ اللازمة لحساب :  $b_0, b_1$

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = (11.8667)(3.8) + (16.8667)(2.8) + \dots + (-3.13333)(-2.2) \\ = 197.6 ,$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 = (11.8667)^2 + (16.8667)^2 + \dots + (-3.1333)^2 = 1019.73 .$$

الآن بالتعويض عن تلك القيم في الصيغتين (٨-١)، (٩-١) نحصل على :

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ = \frac{197.6}{1019.73} = 0.193776 .$$

$$b_0 = \bar{y} - b_1 \bar{x} .$$

$$= 5.2 - (0.193776)(88.133) = -11.8781$$

وعلى ذلك تقديرات المربعات الصغرى للمعلمتين  $\beta_0, \beta_1$  هي :

$$b_0 = -11.8781 \quad , \quad b_1 = 0.193776 .$$

معادلة الانحدار المقطرة سوف تكون:

$$y = -11.8781 + 0.193776x .$$

هناك صيغة جبرية مكافئة للصيغة (٨-١) وذلك لحساب  $b_1$  لا تشمل على انحرافات حول المتوسط ومناسبة لاستخدام الآلة الحاسبة وهي :



$$b_1 = \frac{SXY}{SXX} \quad (١٠-١)$$

حيث:

$$SXX = \sum x_i^2 - \frac{(\sum x_i)^2}{n},$$

$$SXY = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}.$$

الكمية  $\sum x_i^2$  تسمى مجموع المربعات الغير مصححة uncorrected sum of squares لقيم  $x$  و  $\sum x_i^2 / n$  تسمى التصحيح للمتوسط والفرق بينهما يسمى مجموع المربعات المصحح لقيم  $x_i$  حيث  $i = 1, 2, \dots, n$  ، أيضا  $\sum x_i y_i$  يسمى مجموع حاصل الضرب غير المصحح uncorrected sum of products و  $\sum x_i \sum y_i / n$  يسمى التصحيح للمتوسطات. الفرق بينهما يسمى مجموع حاصل الضرب المصحح corrected sum of products لقيم  $x_i, y_i$  و  $i = 1, 2, \dots, n$  . للمشاهدات المعطاة في الجدول (٣-١) والخاصة بالمثال (٣-١) أوجد معادلة الانحدار المقدرة باستخدام الصيغة (١٠-١) لحساب  $b_1$  والصيغة (٩-١) لحساب  $b_0$  . سوف نحسب القيم التالية:

$$\begin{aligned} SXY &= \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} \\ &= 7072 - \frac{(1322)(78)}{15} = 197.6, \end{aligned}$$

$$\begin{aligned} SXX &= \sum x_i^2 - \frac{(\sum x_i)^2}{n} \\ &= 117532 - \frac{(1322)^2}{15} = 1019.73, \end{aligned}$$

$$b_1 = \frac{SXY}{SXX} = \frac{197.6}{1019.73} = 0.193776,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 5.2 - (0.193776)(88.1333) = -11.8781.$$

يتضح من حساب قيمة  $b_1$  من المعادلة (١٠-١) و  $b_0$  من المعادلة (٩-١) أنها نفس القيم التي تم الحصول عليها باستخدام حل المعادلات في (٤-١) و (٥-١) أدنا ولكن الطريقة الأخيرة تعتبر الأسهل.

وعلى ذلك معادلة الانحدار المقدرة للمثال (٣-١) تكون على الشكل التالي :

$$\hat{y} = -11.8781 - 0.193776 x .$$

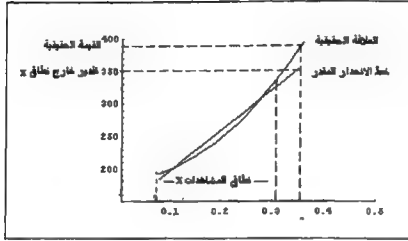
وتسمى معادلة انحدار  $y$  على  $x$  وعلى ذلك عندما ننفق £ 75000 على الإعلان فإننا نرغب في التنبؤ بالمبيعات وذلك بوضع  $x=75$  (بالآلف) في معادلة الانحدار أي أن :

$$\hat{y} = -11.8781 + (75)(0.1938) = 2.6569 = £ 2656900 .$$

ماذا يعني هذا التنبؤ ؟ من الواضح أن هذا لا يعطي أنه في كل مرة ننفق 75000 £ على الإعلان سوف نبيع بالضبط £2656900 في الحقيقة فإن التقدير للمبيعات يمثل قيمة متوسطة. عندما يتم إيجاد معادلة الانحدار المقدرة من القيم المشاهدة لـ  $x$  والتي تتراوح بين 70.000 £ و 105000 £ ثم نستخدم المعادلة المقدرة في حساب مستوى المبيعات الناتج من إنفاق إعلانات قيمتها £75000. في هذه الحالة فإن القيمة للمتغير المستقل في هذه الحالة تقع في مدى القيم المشاهدة وتسمى العملية في هذه الحالة تقع بين interpolation أو placing between.) السؤال الآن ماذا عن القيم المقدرة لمبيعات من إنفاق على الإعلانات يساوي £ 120000 :

$$\hat{y} = -11.8781 + (0.1938)(120) = 11377900 .$$

هنا استخدمنا قيمة لـ  $x$  خارج مدى القيم المشاهدة. تسمى العملية في هذه الحالة (تقع خارج placing أو extrapolated). كلا التقديرين يتعرضان لخطأ ولكن التقدير الذي يقع خارج مدى القيم المشاهدة يكون أقل كفاءة من الذي يقع داخل مدى القيم المشاهدة. هذا يرجع لأن داخل مدى القيم المشاهدة من  $x$  فإننا نعرف سلوك البيانات وكيف يمكن توفيق الخط المستقيم، أما خارج مدى المشاهدات فلا نعرف سلوك البيانات وفي هذه الحالة قد لا يكون الخط المستقيم توفيق جيد لتلك القيم من  $x$ . والمثال على ذلك موضح في شكل (١٥-١) والذي يجعلنا نتخذ الحذر عند الحصول على تقديرات خارج المدى لقيم  $x$ .



شكل (١-١٥)

يوضح جدول (١-٥) القيم المشاهدة  $y$  والقيم التوقعية  $\hat{y}$  والبولقي  $y - \hat{y}$ .

جدول (١-٥)

x	y	$\hat{y}$	$y - \hat{y}$
100	9	7.49283	1.50052
105	8	8.46836	-0.468358
90	5	5.56172	-0.561715
80	2	3.62395	-1.62395
80	4	3.62395	0.376046
85	6	4.59283	1.40717
87	4	4.98039	-0.980387
92	7	5.94927	1.05073
90	6	5.56172	0.438285
95	7	6.5306	0.469404
93	5	6.14304	-1.14304
85	5	4.59283	0.407165
85	4	4.59283	-0.592835
70	3	1.68619	1.31381
85	3	4.59283	-1.59283
1322	78	78	$6.03961 \cdot 10^{-24}$

(٧-١) خواص مقدرات المربعات الصغرى

Properties of the least squares estimators

سوف نوجد المتوسط والتباين لمقدرات المربعات الصغرى للمعلمتين  $\beta_0, \beta_1$  وذلك بالاستفادة من الفروض التي وضعناها على  $\varepsilon_i$  في النموذج  $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$  حيث  $i = 1, 2, \dots, n$ . من المهم أن نتذكر أن قيمتي  $b_0, b_1$  والتي تعتمد على عينة عشوائية معطاة من الحجم  $n$  من المشاهدات هما تقديرين للمعلمتين الحقيقيتين  $\beta_0, \beta_1$ . بفرض تكرار التجربة عدة مرات وفي كل مرة نستخدم نفس القيم لـ  $x$  فإن التقديرين الناتجين  $b_0, b_1$  سوف يختلفان في العادة من تجربة إلى أخرى. تلك الاختلافات تعني أن التقديرات التي نحصل عليها من عينة عشوائية معطاة تمثل قيم لمغيرات عشوائية  $B_0, B_1$ . وبما أن قيم  $x$  تظل ثابتة فإن قيم  $B_0, B_1$  تعتمد على الاختلافات في قيم  $Y$  أو بصورة أدق تعتمد على قيم المتغيرات العشوائية  $Y_1, Y_2, \dots, Y_n$ . الفروض الموضوعة على توزيعات  $\varepsilon_i$  تعني أن  $Y_i$  حيث  $i = 1, 2, \dots, n$  تمثل متغيرات عشوائية غير مرتبطة بمتوسط  $\mu_{Y|x} = \beta_0 + \beta_1 x$  وتباين علم  $\sigma^2$ . أي أن:

$$\sigma_{Y|x_i}^2 = \sigma^2, \quad i = 1, 2, \dots, n.$$

في الجزء التالي سوف نثبت أن المقدرين  $B_0, B_1$  غير متحيزين للمعلمتين  $\beta_0, \beta_1$  على التوالي. وأيضاً سوف نوجد التباين لكل من المقدرين  $B_0, B_1$  وذلك للاستفادة من تباينهما في الحصول على فترات ثقة واختبارات فروض تخصص للمعلمتين  $\beta_0, \beta_1$ .  
بما أن المقدر :

$$B_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (11-1)$$

وذلك لأن :

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}) &= \sum_{i=1}^n (x_i - \bar{x})Y_i - \sum_{i=1}^n (x_i - \bar{x})\bar{Y} \\ &= \sum_{i=1}^n (x_i - \bar{x})Y_i - \bar{Y} \sum_{i=1}^n (x_i - \bar{x}) \end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^n (x_i - \bar{x}) Y_i - \bar{Y}(0) \\ &= \sum_{i=1}^n (x_i - \bar{x}) Y_i . \end{aligned}$$

يمكن كتابة (١١-١) على شكل تركيبة خطية في المتغيرات  $Y_1, Y_2, \dots, Y_n$  كالآتي:

$$B_1 = \sum_{i=1}^n c_i Y_i$$

حيث :

$$c_i = \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2} , \quad i = 1, 2, \dots, n .$$

يكون من السهل إثبات أن  $\sum_{i=1}^n c_i = 0$  . أيضا فإن :

$$\begin{aligned} \sum_{i=1}^n c_i x_i &= \sum_{i=1}^n c_i x_i - \bar{x} \sum_{i=1}^n c_i \\ &= \sum_{i=1}^n c_i (x_i - \bar{x}) = 1 . \end{aligned}$$

وعلى ذلك :

$$\begin{aligned} \mu_{B_1} &= E(B_1) = E\left(\sum_{i=1}^n c_i Y_i\right) \\ &= \sum_{i=1}^n c_i E(Y_i) \\ &= \sum_{i=1}^n c_i (\beta_0 + \beta_1 x_i) \\ &= \beta_0 \sum_{i=1}^n c_i + \beta_1 \sum_{i=1}^n c_i x_i \\ &= \beta_0(0) + \beta_1(1) \\ &= \beta_1 . \end{aligned}$$

تباين المقدّر  $B_1$  هو :

$$\sigma_{B_1}^2 = \text{Var}(B_1) = \text{Var}\left(\sum_{i=1}^n c_i Y_i\right) = \sum_{i=1}^n c_i^2 \text{Var}(Y_i) .$$

وذلك لأن  $Y_i$  متغيرات غير مرتبطة وعلى ذلك فإن تباين المجموع هو بالضبط مجموع التباينات. التباين لكل حد في المجموع هو  $c_i^2 \text{Var}(Y_i)$  ونحت فرض أن  $\text{Var}(Y_i) = \sigma^2$  فإن :

$$\begin{aligned} \text{Var}(B_1) &= \sigma^2 \sum_{i=1}^n c_i^2 = \sigma^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\left[ \sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} \\ &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} . \end{aligned}$$

لإثبات أن المقدّر  $B_0$  غير متحيز للمعلمة  $\beta_0$  فإننا نكتب المقدّر  $B_0$  على الشكل التالي :

$$\begin{aligned} B_0 &= n^{-1} \sum_{i=1}^n Y_i - \bar{x} \sum_{i=1}^n c_i Y_i \\ &= \sum_{i=1}^n (n^{-1} - \bar{x} c_i) Y_i \end{aligned}$$

حيث :

$$(n^{-1} - \bar{x} c_i)$$

ثوابت معلومة تحمل الصفات التالية :

$$\sum (n^{-1} - \bar{x} c_i) = 1, \sum (n^{-1} - \bar{x} c_i) x_i = 0$$

وعلى ذلك :

$$\begin{aligned} \mu_{B_0} &= E(B_0) = \sum (n^{-1} - \bar{x} c_i) (\beta_0 + \beta_1 x_i) \\ &= \beta_0 \sum (n^{-1} - \bar{x} c_i) + \beta_1 \sum (n^{-1} - \bar{x} c_i) x_i \\ &= \beta_0 (1) + B_1(0) = \beta_0 . \end{aligned}$$

ولإيجاد تبين  $B_0$  نتبع الآتي:

$$\begin{aligned}\sigma_{B_0}^2 &= \text{Var}(B_0) = \sum_{i=1}^n [n^{-1} - \bar{x}c_i]^2 \text{Var}(Y_i) \\ &= \sigma^2 \sum_{i=1}^n [n^{-2} - 2n^{-1}\bar{x}c_i + \bar{x}^2 c_i^2] \\ &= \sigma^2 \left[ n^{-1} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] = \sigma^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right].\end{aligned}$$

التغاير بين  $B_0, B_1$  يعرف كالآتي :

$$\text{Cov}(B_0, B_1) = E(B_0 - \beta_0)(B_1 - \beta_1).$$

لإثبات أن :

$$\text{Cov}(B_0, B_1) = -\bar{x} \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

سوف نستخدم على النظرية التالية :

نظرية (١-١)

إذا كان  $a_i, d_i$  ثوابت وكان  $a, d$  متغيرين عشوائيين حيث:

$$a = a_1 Y_1 + a_2 Y_2 + \dots + a_n Y_n,$$

$$d = d_1 Y_1 + d_2 Y_2 + \dots + d_n Y_n$$

و إذا كان  $Y_i, Y_j$  غير مرتبطين حيث  $i \neq j$  وإذا كان  $\text{Var}(Y_i) = \sigma^2$  لكل  $i$  فإن :

$$\text{Cov}(a, d) = (a_1 d_1 + a_2 d_2 + \dots + a_n d_n) \sigma^2.$$

بوضع  $a = B_1$  و  $d = B_0$  وهذا يعني أن:

$$d_i = c_i = \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad a_i = (n^{-1} - \bar{x}c_i)$$

وعلى ذلك :

$$\begin{aligned} \text{Cov}(B_0, B_1) &= \left[ \Sigma(n^{-1} - \bar{x}c_i) \frac{x_i - \bar{x}}{\Sigma(x_i - \bar{x})^2} \right] \sigma^2 \\ &= \frac{\sigma}{n} \frac{\Sigma(x_i - \bar{x})}{\Sigma(x_i - \bar{x})^2} - \bar{x} \frac{\Sigma(x_i - \bar{x})^2}{[\Sigma(x_i - \bar{x})^2]^2} \sigma^2 \\ &= 0 - \bar{x} \frac{\sigma^2}{\Sigma(x_i - \bar{x})^2}. \end{aligned}$$

إن جودة مقدرات المربعات الصغرى  $B_0, B_1$  تنص عليها النظرية التالية :

#### نظرية (١-٢)

تنص هذه النظرية على أنه لنموذج الانحدار الخطي في (١-١) وتحته الفروض  $E(\varepsilon_i) = 0$  و  $\text{Var}(\varepsilon_i) = \sigma^2$  حيث  $E(\varepsilon_i \varepsilon_j) = 0$  فإن مقدرات المربعات الصغرى غير متحيزة ولها أقل تباين وذلك عند مقارنتها بكل المقدرات الأخرى الغير متحيزة والتي يعبر عنها على شكل تركيبة خطية في المتغيرات  $Y_1, Y_2, \dots, Y_n$ . عادة يقال إن مقدرات المربعات الصغرى هي أفضل المقدرات الغير متحيزة الخطية حيث كلمة أفضل تعني أن لها أقل تباين.

لإثبات أقل تباين للمقدر  $B_1$  نتبع الآتي :

فترض أن  $W$  مقدر آخر للمعلمة  $\beta_1$  على الشكل :

$$W = \Sigma k_i Y_i \quad (١٢-١)$$

حيث  $k_i$  لوزان مرجحة جديدة. يمكن كتابة (١٢-١) كما يلي :

$$\begin{aligned} W &= \Sigma k_i (\beta_0 + \beta_1 x_i + \varepsilon_i) \\ &= \beta_0 \Sigma k_i + \beta_1 \Sigma k_i x_i + \Sigma k_i \varepsilon_i, \end{aligned}$$

حيث  $W$  مقدر غير متحيز للمعلمة  $\beta_1$  إذا تحقق الشرطان التاليان :

$$\Sigma k_i = 0, \quad \Sigma k_i x_i = 1.$$



الآن يمكن كتابة التباين للمقدر W كما يلي :

$$\text{Var}(W) = \text{Var}(\sum k_i Y_i) = \sum k_i^2 \text{Var}(Y_i)$$

وبما أن :

$$\text{Var}(\varepsilon_i) = \text{Var}(Y_i) = \sigma^2,$$

فإن :

$$\text{Var}(W) = \sigma^2 \sum k_i^2.$$

بإضافة وطرح القيمة  $\frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$  للحد  $k_i$  في المعادلة السابقة نحصل على :

$$\begin{aligned} \text{Var}(W) &= \sigma^2 \left[ \sum_{i=1}^n \left( k_i - \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2} + \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2} \right)^2 \right] \\ &= \sigma^2 \left[ \sum \left( k_i - \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right)^2 + \frac{(x_i - \bar{x})^2}{(\sum (x_i - \bar{x})^2)^2} + 2 \sum \left( k_i - \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right) \left( \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right) \right] \\ &= \sigma^2 \sum \left( k_i - \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right)^2 + \frac{\sigma^2}{\sum (x_i - \bar{x})^2}, \quad (١٣-١) \end{aligned}$$

من المعادلة (١٣-١) فإن تباين المقدر W سوف يساوي تباين المقدر  $B_1$  إذا تساوت الأوزان  $k_i, c_i$ . لما إذا اختلفت الأوزان فيبقى تباين المقدر  $B_1$  أقل من تباين المقدر W. وبالتالي تعتبر مقدرات المربعات الصغرى مقدرات لها أقل تباين.

#### خواص خط الانحدار المقدر بطريقة المربعات الصغرى

١. مجموع البواقي في أي نموذج انحدار يحتوي على الجزء المقطوع  $\beta_0$  دائماً يساوي صفر، أي أن:

$$\sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n \varepsilon_i = 0.$$

ويمكن إثبات ذلك كما يلي :

$$\begin{aligned}\sum e_i &= \sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)] \\ &= \sum_{i=1}^n [y_i - (\bar{y} - b_1 \bar{x} + b_1 x_i)] \\ &= \sum_{i=1}^n [(y_i - \bar{y}) - b_1 (x_i - \bar{x})] \\ &= \sum_{i=1}^n (y_i - \bar{y}) - b_1 \sum_{i=1}^n (x_i - \bar{x}) = 0.\end{aligned}$$

وهذا يعني أن مجموع القيم المشاهدة  $y_i$  يساوي مجموع القيم التقديرية المقابلة  $\hat{y}_i$  أي أن :

$$\sum y_i = \sum \hat{y}_i .$$

وعادة في العمليات الحسابية قيمة  $\sum e_i$  لا تساوي صفر ولكن قريبة من الصفر حيث تحدث أخطاء نتيجة لتكوين الأرقام العشرية لأي مشاهدة مما يجعل مجموع البواقي غير مساوي للصفر تماما. ويمكن التحقق من ذلك من المشاهدات المعطاة في جدول (٥-١) والخاصة بمثال (٣-١) حيث  $\sum (y_i - \hat{y}_i) \neq 0$  معطاة في العمود الأخير من جدول (٥-١).

٢. خط الحدار المربعات الصغرى المقدر دائما يمرر خلال النقطة  $(\bar{x}, \bar{y})$  ويمكن التحقق من ذلك بقسمة المعادلة (٤-١) على  $n$  نحصل على :

$$\bar{y} = b_0 + b_1 \bar{x} .$$

٣. مجموع البواقي المرجحة بالقيم المقابلة للمتغير المستقل دائما تساوي صفر، أي أن :

$$\sum_{i=1}^n e_i x_i = 0 .$$

ويمكن التحقق من ذلك كما يلي :

$$\begin{aligned}\sum_{i=1}^n e_i x_i &= \sum [(y_i - \bar{y}) - b_1(x_i - \bar{x})]x_i \\ &= \sum (y_i - \bar{y})x_i - b_1 \sum (x_i - \bar{x})x_i \\ &= \sum (y_i - \bar{y})(x_i - \bar{x}) - b_1 \sum (x_i - \bar{x})^2 .\end{aligned}$$

وبالتعويض عن  $b_1$  نحصل على :

$$\sum_{i=1}^n e_i x_i = 0 .$$

٤ . مجموع البواقي المرجحة بالمقيم المقابلة للقيم المقدرة  $\hat{y}_i$  دائماً يساوي الصفر، أي أن :

$$\sum_{i=1}^n e_i \hat{y}_i = 0 .$$

ويمكن التحقق من ذلك كما يلي :

$$\sum_{i=1}^n e_i \hat{y}_i = \sum_{i=1}^n (y_i - \hat{y}_i) \hat{y}_i = \sum_{i=1}^n y_i \hat{y}_i - \sum_{i=1}^n \hat{y}_i^2 ,$$

$$\begin{aligned}\sum_{i=1}^n y_i \hat{y}_i &= \sum_{i=1}^n y_i (\bar{y} + b_1(x_i - \bar{x})) = \bar{y} \sum_{i=1}^n y_i + b_1 \sum_{i=1}^n (x_i - \bar{x}) y_i \\ &= n\bar{y}^2 + b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 ,\end{aligned}$$

$$\begin{aligned}\sum (\hat{y}_i)^2 &= \sum [\bar{y} + b_1(x_i - \bar{x})]^2 \\ &= n\bar{y}^2 + 2b_1\bar{y} \sum (x_i - \bar{x}) + b_1^2 \sum (x_i - \bar{x})^2 \\ &= n\bar{y}^2 + b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \sum y_i \hat{y}_i .\end{aligned}$$

أي أن :

$$\sum_{i=1}^n y_i \hat{y}_i = \sum_{i=1}^n (\hat{y}_i)^2 .$$

وبالتالي فإن :

$$\sum_{i=1}^n e_i \hat{y}_i = 0.$$

### (٨-١) صيغة بديلة للنموذج

#### An alternate form of the Model

هناك صيغة بديلة للنموذج (١-١) أثبتت فائدتها. بفرض أننا أعدنا تعريف المتغير المستقل  $x_i$  على شكل انحرافات عن الوسط الحسابي ، ليكن  $(x_i - \bar{x})$  . وعلى ذلك فإن نموذج الانحدار يصبح :

$$\begin{aligned} Y_i &= \beta_0 + \beta_1(x_i - \bar{x}) + \beta_1\bar{x} + \varepsilon_i \\ &= (\beta_0 + \beta_1\bar{x}) + \beta_1(x_i - \bar{x}) + \varepsilon_i \\ &= \beta'_0 + \beta_1(x_i - \bar{x}) + \varepsilon_i . \end{aligned}$$

لجعل قيم  $\hat{y}_i$  واحدة في كل من النموذج الأصلي والنموذج المحول يكون من الضروري تعديل الجزء المقطوع الأصلي. العلاقة بين الجزء المقطوع الأصلي والمحول هي :

$$\beta'_0 = \beta_0 + \beta_1\bar{x} .$$

المعادلات الطبيعية للمربعات الصغرى لهذا النوع من النماذج تكون على الشكل التالي :

$$nb'_0 = \sum_{i=1}^n y_i ,$$

$$b_1 \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n y_i (x_i - \bar{x}) .$$

وعلى ذلك فإن تقديرات المربعات الصغرى تكون :

$$b'_0 = \bar{y} , \quad b_1 = \frac{\sum_{i=1}^n y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{SXY}{SXX} .$$

وعلى ذلك فإن الجزء المقطوع من هذا النموذج يقدر بـ  $\bar{y}$  أما الميل فلا يتأثر بالتحويل.

هناك عدة مميزات ترتبط بهذا النموذج المحول :

أولاً: المعادلات الطبيعية تكون أسهل في الحل من المعادلات في  $(1-4)$ ،  $(1-5)$ .

ثانياً: المقدرات  $B'_0, B_1 = \frac{\sum XY}{\sum XX}$  غير مرتبطة، أي أن  $\text{Cov} = (B'_0, B_1) = 0$ .

وأخيراً فإن معادلة الانحدار المقطرة:

$$\hat{y} = \bar{y} + b_1(x - \bar{x}) \quad (1-14)$$

والمعادلة المقطرة :

$$\hat{y} = b_0 + b_1 x$$

متكافئتان (أي أن كلاهما يعطي نفس قيم  $\hat{y}$  لنفس القيمة من  $x$ ). ويجب أن يتذكر الباحث أن  $(1-14)$  صحيحة في مجال  $x$  للبيانات الأصلية.

الباقى  $e_i$  سوف يكون :

$$e_i = y_i - \hat{y}_i = y_i - \bar{y} - b_1(x_i - \bar{x})$$

وعلى ذلك :

$$\sum_{i=1}^n e_i = \sum_{i=1}^n (y_i - \bar{y}) - b_1 \sum_{i=1}^n (x_i - \bar{x}) = 0.$$

مثال (1-4)

إذا عرف أن هناك علاقة بين فترة الإصابة بمرض معين وعدد البكتيريا في العضو المصاب، وتم اختيار  $n$  مصاب بهذا المرض وسجلت أطوال فترات إصابتهم بالمرض منذ بدء دخولهم المستشفى يتم الحصول على البيانات المعطاة في جدول (1-6). المطلوب إيجاد معادلة الانحدار المقطرة

$$\hat{y} = b_0 + b_1 x \quad \text{أ- على الشكل :}$$

$$\hat{y} = \bar{y} + b_1(x - \bar{x}) \quad \text{على الشكل :}$$

جدول (٦-١)

x	y	x <sup>2</sup>	xy
9	12	81	108
10	11	100	110
3	8	25	40
7	9	49	63
10	13	100	130
6	10	36	60
7	14	49	98
4	8	16	32
8	11	64	88
6	7	36	42
72	103	556	771

الحل

$$n=10,$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{72}{10} = 7.2 ,$$

-1

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{103}{10} = 10.3 ,$$

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i / n}{\sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 / n}$$

$$= \frac{771 - (72)(103)/10}{556 - (72)^2 / 10}$$

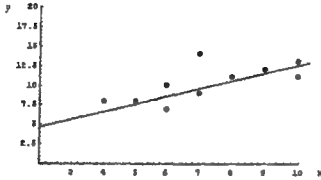
$$= \frac{29.4}{37.6} = 0.781915 ,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 10.3 - (0.781915)(7.2) = 4.67021 .$$

معادلة الانحدار المقترنة سوف تكون على الشكل:

$$\hat{y} = 4.67021 + 0.781915x.$$

والممثلة بيانياً في شكل (١٦-١) مع شكل الانتشار.



شكل (١٦-١)

ب- يمكن الحصول على معادلة الانحدار المقترنة على الصورة :

$$\hat{y} = \bar{y} + b_1(x - \bar{x})$$

كالتالي :

$$\hat{y} = 10.3 + 0.781915(x - 7.2) \quad (١٥-١)$$

وسوف يستخدم النموذجين في أ- و، ب - بالتبادل وفق ما تمليه المناسبة.

(١-٩) تقدير  $\sigma^2$

**Estimation  $\sigma^2$**

لكي نكون قادرين على عمل استدلالات تخص  $\beta_0, \beta_1$  يكون من الضروري الوصول إلى تقدير للمعلمة  $\sigma^2$  والتي ظهرت في الصيغتين السابقتين للتباين الخاص بكل من المقدرين  $B_0, B_1$ . المعلمة  $\sigma^2$  والتي تمثل تباين خطأ النموذج تعكس الاختلاف العشوائي أو اختلاف خطأ التجربة حول خط الانحدار. في الحقيقة يفضل الحصول على تقدير للمعلمة  $\sigma^2$  لا يعتمد على النموذج. عموماً مثل هذا التقدير يكن الحصول عليه فقط في فئات البيانات التي تحتوي على قيم متكررة لـ  $y$  وذلك عند كل قيمة من قيم  $x$  أو عند توفر بعض المعلومات القبلية. عندما لا تتوفر مثل هذه الإمكانيات فإن التقدير للمعلمة  $\sigma^2$  يمكن الحصول عليه بالاعتماد على مجموع مربعات البواقي SSE حيث :

$$\begin{aligned} SSE &= \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 = \sum_{i=1}^n [(y_i - \bar{y}) - b_1(x_i - \bar{x})]^2 \\ &= \sum_{i=1}^n \left[ (y_i - \bar{y})^2 - 2b_1 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) + b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \right] \\ &= SYY - 2b_1 SXY + b_1^2 SXX \\ &= SYY - b_1^2 SXX \end{aligned} \quad (١٦-١)$$

حيث :

$$\begin{aligned} SXX &= \sum_{i=1}^n (x_i - \bar{x})^2 , \\ SXY &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) , \\ SYY &= \sum_{i=1}^n (y_i - \bar{y})^2 . \end{aligned}$$

المعادلة (١٦-١) يمكن كتابتها على الصورة التالية :

$$SSE = SYY - b_1^2 SXX . \quad (١٧-١)$$



الصيغة (١٦-١) والصيغة (١٧-١) تم الحصول عليهما من الحقيقة أن :

$$b_1 = \frac{SXY}{SXX} .$$

مجموع مربعات البواقي له  $n-2$  درجات حرية وذلك لأن درجتين حرية مرتبطتين بتقدير  $b_0, b_1$  في عملية تقدير  $\hat{Y}_i$ .

نظرية (٣-١)

المقدر الغير متحيز للمعلمة  $\sigma^2$  هو :

$$S^2 = \frac{SSE}{n-2} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2} = MSE$$

البرهان

$$E(S)^2 = E\left(\frac{SSE}{n-2}\right) = \frac{E(SSE)}{n-2}$$

ومن (١٦-١) فإن :

$$E(SSE) = E(SYY - B_1^2 SXX)$$

حيث  $B_1$  هو المقدر للمعلمة  $\beta_1$  . وعلى ذلك :

$$\begin{aligned} E(SSE) &= E\left(SYY - B_1^2 \sum_{i=1}^n (x_i - \bar{x})^2\right) \\ &= \sum_{i=1}^n E(Y_i)^2 - nE(\bar{Y}^2) - \sum_{i=1}^n (x_i - \bar{x})^2 E(B_1^2) . \end{aligned}$$

وبما أن :

$$\bar{Y} = B_0 + B_1 \bar{x} + \bar{\varepsilon} ,$$

$$E(Y_i^2) = \text{Var}(Y_i) + [E(Y_i)]^2 = \sigma^2 + (\beta_0 + \beta_1 x_i)^2 ,$$

$$E(\bar{Y}^2) = \text{Var}(\bar{Y}) + [E(\bar{Y})]^2 = \frac{\sigma^2}{n} + (\beta_0 + \beta_1 \bar{x})^2 ,$$

$$E(B_1^2) = \text{Var}(B_1) + [E(B_1)]^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} + \beta_1^2 .$$

وعلى ذلك فإن :

$$E(SSE) = \sum_{i=1}^n [\sigma^2 + (\beta_0 + \beta_1 x_i)^2] - n \left[ \frac{\sigma^2}{n} + (\beta_0 + \beta_1 \bar{x})^2 \right]$$

$$- \sum_{i=1}^n (x_i - \bar{x})^2 \left[ \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} + \beta_1^2 \right]$$

$$= n\sigma^2 + \sum_{i=1}^n (\beta_0 + \beta_1 x_i)^2 - \sigma^2 - n(\beta_0 + \beta_1 \bar{x})^2 - \sigma^2$$

$$- \beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 .$$

أي أن :

$$E(SSE) = (n-2)\sigma^2 + \sum_{i=1}^n (\beta_0 + \beta_1 x_i)^2 - n(\beta_0 + \beta_1 \bar{x})^2 - \beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 .$$

وبما أن مجموع الحدود الثلاثة الأخيرة يساوي الصفر لذلك فإن :

$$E(SSE) = (n-2)\sigma^2 .$$

وعلى ذلك :

$$E(S^2) = \frac{E(SSE)}{n-2} = \frac{(n-2)\sigma^2}{n-2} = \sigma^2 .$$

أي أن  $S^2$  مقدر غير متحيز للمعلمة  $\sigma^2$  .

سوف نرمز لقيمة من قيم الإحصاء  $S^2$  بالرمز  $s^2$  حيث:

$$s^2 = \frac{\sum (y_i - \hat{y}_j)^2}{n - 2}$$

ويجب الأخذ في الاعتبار أن هذا التقدير يعتمد على النموذج.

متوسط مربع الخطأ أو متوسط مربعات البواقي

### Error mean square or residual mean squares

الجنر التربيعي لـ  $S^2$  يسمى في بعض الأحيان الخطأ القياسي standard error للانحدار وله نفس وحدات القياس مثل قيم المتغير التابع  $Y$ . ولأن  $S^2$  يعتمد على مجموع مربعات البواقي فعند عدم تحقق أي من الفروض الخاصة بالأخطاء في النموذج أو في حالة عدم الاختيار الصحيح للنموذج فإن ذلك يجعل  $S^2$  مقدر غير جيد للمعلمة  $\sigma^2$ .

مثال (٥-١)

يعطي جدول (٧-١) أوزان وأطوال مجموعة من الذكور البالغين والمطلوب إيجاد معادلة خط الانحدار المقدرة، وإيجاد تقدير لـ  $\sigma^2$ .

جدول (٧-١)

x	y	$x^2$	xy
159	68	25281	10812
180	88	32400	15840
175	79	30625	13825
150	65	22500	9750
170	70	28900	11900
171	73	29241	12483
165	63	27225	10395
176	74	30976	13024
1346	580	227148	98029

الحل

$$n = 8,$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1346}{8} = 168.25 ,$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{580}{8} = 72.5 ,$$

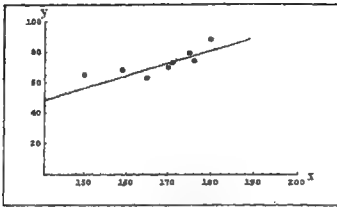
$$\begin{aligned} b_1 = \frac{S_{XY}}{S_{XX}} &= \frac{\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i / n}{\sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 / n} \\ &= \frac{(98029) - (1346)(580)/8}{227148 - (1346)^2 / 8} \\ &= \frac{444}{683.5} = 0.649598 , \end{aligned}$$

$$b_0 = \bar{y} - b_1 \bar{x} = 72.5 - (0.649598)(168.25) = -36.7948 .$$

معادلة الانحدار المقدرة سوف تكون على الشكل:

$$\hat{y} = -36.7948 + 0.649598x .$$

والممثلة بيانياً في شكل (١٧-١) مع شكل الإنتشار .



شكل (١٧-١)

لحساب  $s^2$  نتبع الخطوات التالية:

(١) نحسب  $SY Y$  كالآتي :

$$\begin{aligned} SY Y &= \sum y_i^2 - \frac{(\sum y_i)^2}{n} \\ &= 42508 - \frac{(580)^2}{8} \\ &= 458 . \end{aligned}$$

(٢) نحسب مجموع مربعات الأخطاء (البواقي) من الصيغة التالية :

$$\begin{aligned} SSE &= SY Y - b_1 SXY \\ &= 458 - (0.649598)(444) \\ &= 169.579, \end{aligned}$$

(٣) نحسب التقدير  $s^2$  للمعلمة  $\sigma^2$  من الصيغة التالية :

$$s^2 = \frac{SSE}{n-2} = \frac{169.579}{6} = 28.2631 .$$

### (١٠-١) استدلالات تخص معاملات الانحدار

#### Inferences concerning the regression coefficients

بجانب تقدير العلاقة الخطية بين  $Y, x$  لأغراض التنبؤ فإن القائم على التجربة يهتم بالوصول إلى استدلالات تخص الميل والجزء المقطوع. إن إجراء اختبارات فروض والحصول على فترات ثقة لكل من  $\beta_1, \beta_0$  يحتاج إلى وضع فروض إضافية على نموذج الانحدار (١-١) حيث يفترض أن كل من  $\varepsilon_i$  ، حيث  $i=1, 2, \dots, n$ ، تتبع توزيعاً طبيعياً. هذا الافتراض يعني أن  $Y_1, Y_2, \dots, Y_n$  أيضاً تتبع توزيعات طبيعية حيث  $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ . وبما أن كل من  $\beta_0, \beta_1$  يمثل تركيبة خطية من توزيعات طبيعية مستقلة فإن كل من  $\beta_0, \beta_1$  أيضاً يتبعان توزيعات طبيعية حيث  $B_0 \sim N(\beta_0, \sigma_{\beta_0}^2)$  و  $B_1 \sim N(\beta_1, \sigma_{\beta_1}^2)$  كما تنص النظرية التالية :

#### نظرية (١ - ٤)

إذا كان  $Y_1, Y_2, \dots, Y_n$  تمثل متغيرات عشوائية مستقلة تتبع توزيعات طبيعية بمتوسطات  $\mu_1, \mu_2, \dots, \mu_n$  وتباينات  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$  على التوالي فإن المتغير العشوائي:

$$u = d_1 Y_1 + d_2 Y_2 + \dots + d_n Y_n ,$$

له توزيع طبيعي بمتوسط :

$$\mu_u = d_1 \mu_1 + d_2 \mu_2 + \dots + d_n \mu_n ,$$

وتباين :

$$\sigma_u^2 = d_1^2 \sigma_1^2 + d_2^2 \sigma_2^2 + \dots + d_n^2 \sigma_n^2 .$$

حيث  $d_i$  ثوابت.

وعلى ذلك المتغير العشوائي  $B_1$  ، حيث  $B_1 = \sum c_i Y_i$  ،

$$c_i = d_i , \quad c_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2} ,$$

يتبع توزيعاً طبيعياً بمتوسط :

$$\mu_{B_1} = \beta_1 .$$

وتباين :

$$\sigma_{B_1}^2 = \frac{\sigma^2}{SXX}.$$

وإذا كان:

$$Z = \frac{B_1 - \beta_1}{\sqrt{\frac{\sigma^2}{SXX}}},$$

متغير عشوائي يتبع التوزيع الطبيعي القياسي و  $V = (n-2)S^2/\sigma^2$  متغير عشوائي يتبع توزيع مربع كاي بدرجات حرية  $n-2$  و مستقل عن  $Z$ . فإن :

$$T = \frac{Z}{\sqrt{\frac{V}{n-2}}} = \frac{B_1 - \beta_1}{\sqrt{\frac{S^2}{SXX}}},$$

متغير عشوائي له توزيع  $t$  بدرجات حرية  $n-2$ .

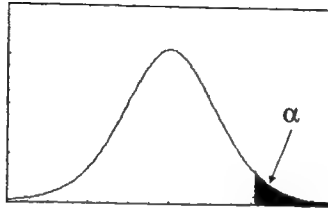
فترة ثقة للمعلمة  $\beta_1$

### Confidence interval for $\beta_1$

سوف نستخدم المتغير  $T$  في إيجاد  $100(1-\alpha)\%$  فترة ثقة للمعلمة  $\beta_1$  على الشكل التالي :

$$b_1 - t_{\alpha/2}(n-2)\sqrt{\frac{s^2}{SXX}} < \beta_1 < b_1 + t_{\alpha/2}(n-2)\sqrt{\frac{s^2}{SXX}}.$$

حيث  $t_{\alpha/2}(n-2)$  تستخرج من جدول توزيع  $t$  في الملحق (١) والتي توجد على المحور الأفقي تحت منحنى توزيع  $t$  بدرجات حرية  $(n-2)$  والتي المساحة على يمينها قدرها  $\alpha/2$  كما هو موضح في شكل (١٨ - ١)



$t_{\alpha/2}$

شكل (١٨-١)

مثال (١-١)

تعتبر كمية الرطوبة في منتج ما لها تأثير على كثافة المنتج النهائي، تم مراقبة المنتج وقياس كثافته و البيانات المسجلة معطاة في جدول (١-٨) في شكل منفردة .

جدول (١-٨)

x	y	$x^2$	xy
4.7	3	22.09	14.1
5	3	25	15
5.2	4	27.04	20.8
5.2	10	27.04	52.
5.9	2	34.81	11.8
4.7	9	22.09	42.3
5.9	3	34.81	17.7
5.2	7	27.04	36.4
5.9	6	34.81	35.4
5.6	6	31.36	33.6
5.	4	25.	20.
58.3	57	311.09	299.1



قدر معالم نموذج الانحدار الخطي البسيط وأوجد 95% فترة ثقة للمعلمة  $\beta_1$ .

الحل

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$= \frac{299.1 - \frac{(58.3)(57)}{11}}{311.09 - \frac{(58.3)^2}{11}}$$

$$= \frac{-3}{2.1} = -1.42857,$$

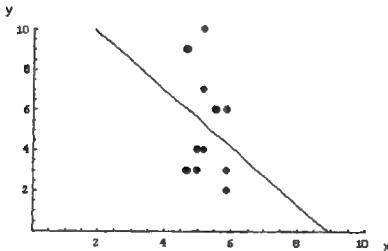
$$b_0 = \bar{y} - b_1 \bar{x} = 5.18182 - (-1.42857)(5.3)$$

$$= 12.7532.$$

معادلة الانحدار المقدرة سوف تكون على الشكل :

$$\hat{Y} = 12.7532 - 1.42857x$$

والممثلة بيانياً في شكل (١-١٩) مع شكل الانتشار .



شكل (١-١٩)

القيم اللازمة لحساب  $s^2$  معطاة في جدول (٩-١) .

جدول (٩-١)

y	$\hat{y}$	y - $\hat{y}$	(y - $\hat{y}$ ) <sup>2</sup>
3	6.03896	- 3.03896	9.23528
3	5.61039	- 2.61039	6.81413
4	5.32468	- 1.32468	1.75476
10	5.32468	4.67532	21.8587
2	4.32468	- 2.32468	5.40412
9	6.03896	2.96104	8.76775
3	4.32468	- 1.32468	1.75476
7	5.32468	1.67532	2.80671
6	4.32468	1.67532	2.80671
6	4.75325	1.24675	1.55439
4	5.61039	- 1.61039	2.59335
57	57	2.66454 × 10 <sup>-13</sup>	65.3506

الآن :

$$s^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2} = \frac{65.3506}{9} = 7.26118 .$$

وباستخدام جدول توزيع t في الملحق (١) فإن  $t_{0.025}(9) = 2.262$  . إذا 95% فترة ثقة للمعلمة  $\beta_1$  تحسب كالآتي :

$$b_1 - t_{\alpha/2}(n-2) \sqrt{\frac{s^2}{SXX}} < \beta_1 < b_1 + t_{\alpha/2}(n-2) \sqrt{\frac{s^2}{SXX}} .$$

أي أن :

$$-1.42857 - 2.262 \sqrt{\frac{7.26118}{2.1}} < \beta_1 < -1.42857 + 2.262 \sqrt{\frac{7.26118}{2.1}} .$$

وعلى ذلك :

$$-1.42857 - 2.262(1.85949) < \beta_1 < -1.42857 + 2.262(1.85949) .$$

والتي تختصر إلى :

$$-5.63503 < \beta_1 < 2.77789 .$$

اختبارات فروض تخصص الميل

Hypothesis testing on the slope

$$H_0 : \beta_1 = \beta_1^*$$

لاختبار فرض العدم

ضد فرض بديل مناسب :

$$H_1 : \beta_1 \neq \beta_1^*$$

أو

$$H_1 : \beta_1 > \beta_1^*$$

أو

$$H_1 : \beta_1 < \beta_1^* .$$

يمكننا استخدام توزيع  $t$  بدرجات حرية  $n-2$  للحصول على منطقة رفض .  
قرارنا سوف يعتمد على القيمة:

$$t = \frac{b_1 - \beta_1^*}{\sqrt{s^2 / s_{XX}}} .$$

الطريقة المستخدمة موضحة من المثال (١-٦) وذلك باستخدام  
قيمة  $b_1 = -1.42857$  وبالاعتماد على البيانات في جدول (٨-١) و جدول (٩-١) .

سوف نختبر فرض العدم أن  $H_0 : \beta_1 = -1.7$  ضد البديل أن  $H_1 : \beta_1 < -1.7$

الحل

$$H_0 : \beta_1 = -1.7$$

ضد الفرض البديل:

$$H_1 : \beta_1 < -1.7 ,$$

$$\alpha = 0.05,$$

$t_{0.05}(9) = 1.833$  (درجته حرية  $n-2 = 11-2 = 9$ ) ومنطقة الرفض  $T < -1.833$

$$t = \frac{-1.42857 - (-1.7)}{\sqrt{\frac{7.26118}{2.1}}} = \frac{0.27143}{1.85999} = 0.14597 .$$

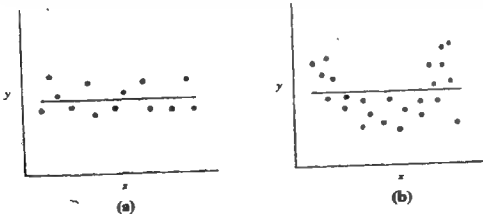
وبما أن قيمة  $t$  المحسوبة تقع في منطقة القبول نقبل  $H_0$  .

حالة خاصة من فرض العدم  $H_0: \beta_1 = \beta_1^*$  هي  $H_0: \beta_1 = 0$

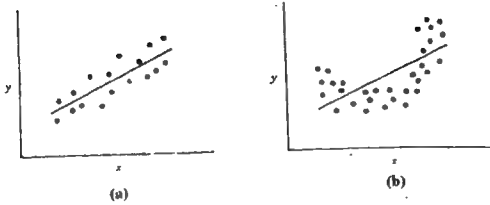
ضد الفرض البديل :

$$H_1: \beta_1 \neq 0 .$$

الفرض السابق يرتبط بمعنوية الانحدار فعند قبول فرض العدم  $H_0: \beta_1 = 0$  فهذا يعني عدم وجود علاقة خطية بين  $x, Y$  . ويجب أن نعلم أن هذا يعني إما أن  $x$  لها قيمة صغيرة في تفسير الاختلاف في  $y$  وأن أفضل تقدير لـ  $y$  عند أي قيمة لـ  $x$  هو  $\hat{y} = \bar{y}$  كما هو موضح في شكل (٢٠-١) أو أن العلاقة الحقيقية بين  $x, Y$  ليست خطية كما هو موضح في شكل (٢٠-١) . أو كبدل وعندما نرفض فرض العدم  $H_0: \beta_1 = 0$  ، فإن هذا يعني أن  $x$  لها قيمة في تفسير الاختلاف في  $y$  . إن رفض  $H_0: \beta_1 = 0$  قد يعني إما أن نموذج الخط المستقيم هو الأنسب كما هو موضح في شكل (٢١-١) أو أن نتائج أفضل يمكن الحصول عليها بإضافة حدود من رتبة عليا من كثيرات الحدود في  $x$  كما هو موضح في شكل (٢١-٢) .



شكل (٢٠-١)



شكل (١-٢١)

باستخدام قيمة  $b_1 = -1.42857$  في المثال (١ - ٦) اختبر فرض العدم  
 أن  $H_0: \beta_1 = 0$  ضد الفرض البديل  $H_1: \beta_1 \neq 0$

الحل

$$H_0: \beta_1 = 0,$$

ضد الفرض البديل

$$H_1: \beta_1 \neq 0,$$

$$\alpha = 0.05,$$

$$T < -2.262 \text{ أو } T > 2.262 \text{ ومنطقة الرفض } t_{0.025}(9) = 2.262$$

$$t = \frac{b_1 - 0}{\sqrt{\frac{s^2}{SXX}}}$$

$$= \frac{-1.42857}{\sqrt{\frac{7.26118}{2.1}}} = \frac{-1.42857}{1.8594} = -0.768259.$$

وبما أن قيمة  $t$  المحسوبة تقع في منطقة القبول نقبل  $H_0$ .

### استدلال إحصائي على الجزء المقطوع

#### Statistical inference on the intercept

أيضا المتغير العشوائي  $B_0$  له توزيع طبيعي بمتوسط :

$$\mu_{B_0} = \beta_0 .$$

وتباين :

$$\sigma_{B_0}^2 = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{SXX} \right) .$$

وبما أن المتغير العشوائي :

$$Z = \frac{\beta_0 - B_0}{\sqrt{\sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{SXX} \right)}} ,$$

يتبع التوزيع الطبيعي القياسي . وحيث أن  $V = (n-2) S^2 / \sigma^2$  متغير عشوائي يتبع توزيع مربع كاي بدرجات حرية  $n-2$  ومستقل عن  $Z$  فإن المتغير العشوائي :

$$T = \frac{Z}{\sqrt{\frac{V}{n-2}}} = \frac{B_0 - \beta_0}{\sqrt{S^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{SSX} \right)}}$$

يتبع توزيع  $t$  بدرجات حرية  $n-2$

### فترة ثقة للمعلمة $\beta_0$

#### Confidence interval for $\beta_0$

سوف يستخدم المتغير  $T$  للحصول على  $100(1-\alpha)\%$  فترة ثقة للمعلمة  $\beta_0$  كالتالي :

$$b_0 - t_{\alpha/2}(n-2) \sqrt{s^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{SXX} \right)} < \beta_0 < b_0 + t_{\alpha/2}(n-2) \sqrt{s^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{SXX} \right)} .$$

والآن لإيجاد 95% فترة ثقة للمعلمة  $\beta_0$  في خط الانحدار  $\mu_{Y|x} = \beta_0 + \beta_1 x$  بالاعتماد على البيانات في جدول (٨-١) و جدول (٩-١) الخاصة بالمثال (٦-١) نتبع الآتي :

$$s^2 = 7.26118 \text{ و } SXX = 2.1 \text{ و } \bar{x} = 5.3$$

$$b_0 = 12.7532 .$$

وعلى ذلك في 95% فترة ثقة للمعلمة  $\beta_0$  تعطى على الشكل :

$$b_0 - t_{\alpha/2}(n-2)s^2\left[\frac{1}{n} + \frac{\bar{x}^2}{SXX}\right] < \beta_0 < b_0 + t_{\alpha/2}(n-2)s^2\left[\frac{1}{n} + \frac{\bar{x}^2}{SXX}\right]$$

وعلى ذلك :

$$12.7532 - 2.262 \sqrt{7.26118 \left( \frac{1}{11} + \frac{5.3^2}{2.1} \right)}$$

$$< \beta_0 < 12.7532 + 2.262 \sqrt{7.26118 \left( \frac{1}{11} + \frac{5.3^2}{2.1} \right)}$$

أي أن :

$$12.7532 - 2.262(9.88873) < \beta_0 < 12.7532 + 2.262(9.88873) .$$

والتي تختزل إلى :

$$-9.61663 < \beta_0 < 35.1231 .$$

**$\beta_0$  اختبارات فروض تخصص**

### Hypothesis testing for $\beta_0$

لاختبار فرض العدم  $H_0: \beta_0 = \beta_0^*$  ضد أي فرض بديل مناسب فإننا مرة

أخرى سوف نستخدم توزيع  $t$  بدرجات حرية  $n-2$  للحصول على منطقة الرفض وبالتالي فإن قرارنا سوف يعتمد على القيمة :

$$t = \frac{b_0 - \beta_0^*}{\sqrt{s^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{SXX} \right)}}$$

الطريقة المتبعة لاختبار فرض العدم موضحة باستخدام بيانات المثال (١ - ٦) عند مستوى معنوي  $\alpha = 0.05$  حيث فرض العدم:

$$H_0 : \beta_0 = 0$$

ضد الفرض البديل

$$H_1 : \beta_0 \neq 0.$$

$$t = \frac{b_0 - 0.0}{\sqrt{s^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{SXX} \right)}} = \frac{12.7532}{\sqrt{7.26118 \left[ \frac{1}{11} + \frac{5.3^2}{2.1} \right]}} = \frac{12.7532}{9.88873} = 1.28967016.$$

$t_{0.025}(9) = 2.262$  ومنطقة الرفض  $T > 2.262$  أو  $T < -2.262$ . وبما أن قيمة  $t$  المحسوبة تقع في منطقة القبول نقبل  $H_0$ .

### (١١-١) التنبؤ

#### Prediction

يمكن استخدام المعادلة  $\hat{y} = b_0 + b_1x$  للتنبؤ بقيمة  $\mu_{Y|x_0}$  حيث  $x$  ليس من الضرورة أن تكون واحدة من  $x_1, x_2, \dots, x_n$  في العينة العشوائية من الحجم  $n$  للملاحظات :  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ . أيضا يمكن استخدام المعادلة  $\hat{y} = b_0 + b_1x$  للتنبؤ بقيمة واحدة جديدة  $\hat{y}_0$  للمتغير  $Y|x_0$ . سوف نتوقع أن خطأ التنبؤ يكون أعلى في حالة قيمة واحدة متنبأ بها عنه في حالة التنبؤ بالمعوسط وهذا سوف يؤثر على طول فترة الثقة للمعالم المراد تقديرها.

فترة ثقة لـ  $\mu_{Y|x_0}$

بفرض أن باحث يرغب في الحصول على فترة ثقة للمعلمة  $\mu_{Y|x_0}$ . سوف نستخدم المقدر  $Y_0 = B_0 + B_1x_0$  لتقدير  $\mu_{Y|x_0} = \beta_0 + \beta_1x_0$ . يمكن إثبات أن  $\hat{Y}_0$  مقدر غير متحيز لـ  $\mu_{Y|x_0}$ . كما أن تباين  $\hat{Y}_0$  هو :



$$\text{Var}(\hat{Y}_0) = \left[ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right] \sigma^2.$$

لإثبات ذلك نضع الآتي :

بما أن :

$$\hat{Y}_0 = B_0 + B_1 x_0 \quad (18-1)$$

وبما أن :

$$B_0 = \bar{Y} - B_1 \bar{x} \quad (19-1)$$

وبالتعويض عن  $B_0$  في (18-1) بقيمتها في (19-1) نحصل على :

$$\hat{Y}_0 = \bar{Y} + B_1(x_0 - \bar{x}).$$

وبأخذ التباين للطرفين نحصل على :

$$\begin{aligned} \text{Var}(\hat{Y}_0) &= \text{Var}(\bar{Y}) + (x_0 - \bar{x})^2 \text{Var}(B_1) + 2(x_0 - \bar{x}) \text{Cov}(\bar{Y}, B_1) \\ &= \text{Var}(\bar{Y}) + (x_0 - \bar{x})^2 \text{Var}(B_1) \\ &= \sigma^2 \left[ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right]. \end{aligned}$$

فيما يلي إثبات أن  $\text{Cov}(\bar{Y}, B_1) = 0$  :

بما أن  $\text{Cov}(\bar{Y}, B_1)$  يعتمد على دالتين خطيتين في  $Y_i$  هما :

$$\bar{Y} = \sum \frac{1}{n} Y_i, B_1 = \sum c_i Y_i.$$

لذلك سوف نعتمد على نظرية (١-١) حيث :

$$a_i = \frac{1}{n}, d_i = c_i, i=1, 2, \dots, n, \text{ لكل } i \text{ و على ذلك :}$$

$$\text{Cov}(\bar{Y}, B_0) = \frac{\sigma^2}{n} \sum c_i = \frac{\sigma^2}{n} (0) = 0$$

و ذلك لأن  $\sum c_i = 0$  حيث :

$$c_i = \frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

كما أوضحنا من قبل.

وعلى ذلك التوزيع العيني للإحصاء  $\hat{Y}_0 - \mu_{Y|x_0}$  يتبع توزيع طبيعي بمتوسط :

$$\mu_{\hat{Y}_0 - \mu_{Y|x_0}} = E(\hat{Y}_0 - \mu_{Y|x_0}) = (\beta_0 + \beta_1 x_0) - (\beta_0 + \beta_1 x_0) = 0$$

وتباين :

$$\sigma_{\hat{Y}_0 - \mu_{Y|x_0}}^2 = \sigma^2 \left[ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right].$$

وعلى ذلك فإن الإحصاء:

$$T = \frac{\hat{Y}_0 - \mu_{Y|x_0}}{\sqrt{S^2 \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right)}}$$

له توزيع  $t$  بدرجات حرية  $n-2$ .

**فترة ثقة للمعلمة  $\mu_{Y|x_0}$**

يمكن الحصول على 100% (1- $\alpha$ ) فترة ثقة للمعلمة  $\mu_{Y|x_0}$  من الصيغة التالية :

$$\hat{y}_0 - t_{\alpha/2}(n-2) \sqrt{s^2 \left[ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right]} < \mu_{Y|x_0} < \hat{y}_0 + t_{\alpha/2}(n-2) \sqrt{s^2 \left[ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right]}.$$

**مثال (٧-١)**

قام باحث بجمع بيانات عن عدد الأقراص الممغنطة المستخدمة (x) وزمن الخدمة (y) بال دقائق لعملاء عددهم 12 والبيانات معطاه في جدول (١٠-١) المطلوب:

(أ) إيجاد معادلة الانحدار المقترنة. (ب) إيجاد 95% فترة ثقة للمعلمة  $\mu_{Y|4}$ .

جدول (١٠-١)

x	y	$x^2$	xy
4	197	16	788
6	272	36	1632
2	100	4	200
5	228	25	1140
7	327	49	2289
6	279	36	1674
3	148	9	444
8	377	64	3016
5	238	25	1190
3	142	9	426
1	66	1	66
5	239	25	1195
55	2613	299	14060

الحل

$$SXY = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} \quad (١)$$

$$= 14060 - \frac{(55)(2613)}{12} = 2083.75,$$

$$SXX = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$= 299 - \frac{(55)^2}{12} = 46.91667,$$

$$b_1 = \frac{SXY}{SXX} = \frac{2083.75}{46.9167} = 44.41385,$$

$$b_0 = \bar{y} - b_1 \bar{x},$$

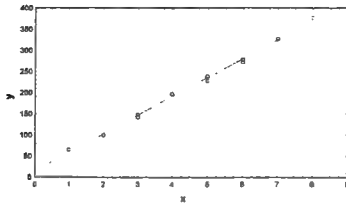
$$= 217.75 - (44.41385)(4.58333)$$

$$= 14.187 .$$

وعلى ذلك فإن معادلة خط الانحدار المقدرة هي :

$$\hat{y} = 14.187 + 44.41385x.$$

والموضحة بيانيا في شكل ( ٢٢-١ ) مع شكل الانتشار.



شكل ( ٢٢-١ )

لحساب SSE نحسب البواقي من جدول (١١-١) .

جدول (١١-١) .

y	$\hat{y}$	y - $\hat{y}$	(y - $\hat{y}$ ) <sup>2</sup>
197	191.842	5.15808	26.6058
272	280.67	-8.66963	75.1624
100	103.014	-3.01421	9.08546
228.	236.256	-8.25577	68.1578
327	325.083	1.91652	3.67304
279	280.67	-1.66963	2.78765
148	147.428	0.571936	0.327111
377	369.497	7.50266	56.29
238	236.256	1.74423	3.04233
142	147.428	-5.42806	29.4639
66	58.6004	7.39964	54.7547
239	236.256	2.74423	7.53078
2613	2613	$-1.49214 \times 10^{-14}$	336.881

وعلى ذلك:

$$s^2 = \frac{SSE}{n-2} = \frac{336.8810}{10} = 33.6881.$$

عندما  $x_0 = 4$  فإن :

$$\begin{aligned}\hat{y}_0 &= 14.187 + (44.41385)(4) \\ &= 191.84\end{aligned}$$

عرفنا مما سبق أن :

$$SXX = 46.91667, \bar{x} = 4.58333, s^2 = 33.6888,$$

$t_{0.025}(10) = 2.228$  . وعلى ذلك 95% فترة ثقة للمعلمة  $\mu_{Y|4}$  هي :

$$191.84 - 2.228 \sqrt{33.6888 \left[ \frac{1}{12} + \frac{(4 - 4.58333)^2}{46.91667} \right]} < \mu_{Y|4} <$$

$$191.84 + 2.228 \sqrt{33.6888 \left[ \frac{1}{12} + \frac{(4 - 4.58333)^2}{46.91667} \right]} .$$

أي أن :

$$191.84 - (2.228)(1.7469) < \mu_{Y|4} < 191.84 + (2.228)(1.7469) .$$

والتي تختصر إلى :

$$187.94791 < \mu_{Y|4} < 195.73209 .$$

الآن وبتكرار العمليات الحسابية السابقة لقيم مختلفة من  $x_0$  يمكن الحصول على فترات الثقة المقابله لكل  $\mu_{Y|x_0}$  كما يتضح من المثال التالي:

#### مثال (١ - ٨)

يعطي جدول (١ - ١٢) متوسط ضربات الخصم ( $\bar{x}$ ) ونسبة الفوز لفريق ما ( $y$ ) وذلك في لعبة كرة الملة .

والمطلوب:

(١) رسم شكل الانتشار مع خط الانحدار المقدر .

(ب) إيجاد 95% فترة ثقة لـ  $\mu_{y|x}$  لعدة قيم من  $x$  ووضحها بيانيا .

جدول (١ - ١٢)

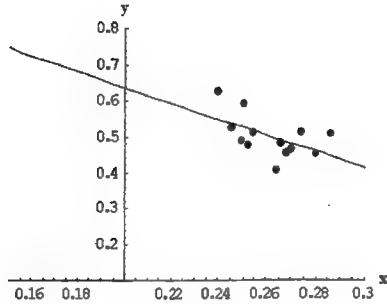
x	y	$x^2$	xy
0.24	0.625	0.0576	0.15
0.254	0.512	0.064516	0.130048
0.249	0.488	0.062001	0.121512
0.245	0.524	0.060025	0.12838
0.25	0.588	0.0625	0.147
0.252	0.475	0.063504	0.1197
0.254	0.513	0.064516	0.130302
0.27	0.463	0.0729	0.12501
0.274	0.512	0.075076	0.140288
0.264	0.405	0.069696	0.10692
0.28	0.45	0.0784	0.126
0.266	0.48	0.070756	0.12768
0.268	0.456	0.071824	0.122208
0.286	0.506	0.081796	0.144716
3.652	6.997	.9551	1.81976

الحل

(أ) شكل الانتشار مع خط الانحدار المقدر موضح في شكل (١ - ٢٣) حيث معادلة الانحدار المقدرة ثم حسابها باستخدام برنامج Mathematica على الحاسب الآلي وكانت :

$$y = 1.07813 - 2.2171x.$$

(ب) يعطي جدول (١ - ١٣) فترات ثقة لـ  $\mu_{y|x}$  وذلك لعدة قيم من  $x$  وتلك الفترات موضحة بيانيا في شكل (١ - ٢٤) وتم للحصول عليها باستخدام الحزم الجاهزة لبرنامج Mathematica . للرمز SE في جدول (١ - ١٣) هو الخطأ القياسي للانحدار المقدر S والرمز CI يرمز لـ 95% فترة ثقة للمعطى  $\beta_1$  .

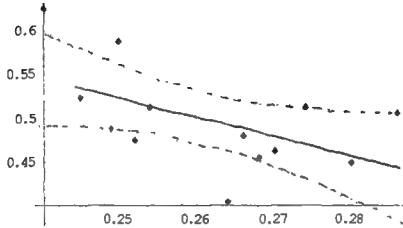


شكل (١-٢٣)

ينضج من شكل (١-٢٤) أن طول فترة الثقة يزداد كلما زادت  $|x - \bar{x}|$ .

جدول (١-١٣)

	المشاهدة	القيم المتنبأ بها	SE	CI
{Mean Prediction CTable→	Observed	Predicted		
	0.625	0.546028	0.0242175	{0.493263,0.598793}
	0.512	0.514989	0.0146246	{0.483124,0.546853}
	0.488	0.526074	0.0174281	{0.488102,0.564047}
	0.524	0.534943	0.0202533	{0.490814,0.579071}
	0.588	0.523857	0.0167906	{0.487273,0.560441}
	0.475	0.519423	0.0156224	{0.485385,0.553461}
	0.513	0.514989	0.0146246	{0.483124,0.546853}
	0.463	0.479515	0.0157797	{0.445134,0.513896}
	0.512	0.470647	0.0182922	{0.430791,0.510502}
	0.405	0.492818	0.0133495	{0.463732,0.521904}
	0.45	0.457344	0.0228174	{0.407629,0.507059}
	0.48	0.488383	0.0139328	{0.458027,0.51874}
	0.456	0.483949	0.0147553	{0.4518,0.516098}
	0.506	0.444042	0.0278533	{0.383354,0.504729}



شكل (٩-٧٤)

#### فترة ثقة تنبؤية لمشاهدة مستقبلية

دفع آخر من الفترات والذي يحدث ليس بينه وبين فترة ثقة لـ  $\mu_Y|x_0$  وهو فترة تنبأ لمشاهدة مستقبلية للاستجابة  $y_0$  عند مستوى معين  $x_0$  للمتغير المستقل. في الحقيقة في كثير من المجالات العلمية يكون الاهتمام بفترة ثقة لمشاهدة مستقبلية أكثر من الاهتمام بفترة ثقة للمتوسط. فعلى سبيل المثال عند دراسة العلاقة بين مبيعات شركة وعدد الأشخاص الذين اعمارهم 16 سنة فما فوق ، استنادا إلى بيانات من السنوات العشرة الماضية ومع توافر إسقاط سكانى موثوق لعدد الأشخاص الذين اعمارهم 16 سنة فأكثر في السنة القادمة ، يرغب الاقتصادى التنبؤ بمبيعات السنة القادمة. عموما الفكرة الأساسية لفترة التنبؤ هي ان توزيع  $Y$  تقع فيه اغلب المشاهدات والادعاء بان المشاهدة الفهم سوف نفع في هذا المدى.

للحصول على  $100\%(1-\alpha)$  فترة ثقة لأي قيمة مفردة  $y_0$  للمتغير  $Y|x_0$  سوف سمد على الإحصاء  $\hat{Y}_0 - Y_0$  . يمكن اثبات ان التوزيع العيني للإحصاء  $\hat{Y}_0 - Y_0$  يتبع توزيعا طبيعيا بمتوسط :

$$\mu_{\hat{Y}_0 - Y_0} = E(\hat{Y}_0 - Y_0) = 0 \quad (٧٠-١)$$

وتباين :



$$\sigma_{\hat{Y}_0 - Y_0}^2 = \sigma^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right] . \quad (٢١-١)$$

لإثبات (٢٠-١) نتبع الاتى :

$$E(\hat{Y}_0 - Y_0) = (\beta_0 + \beta_1 x_0) - (\beta_0 + \beta_1 x_0) = 0 .$$

لإثبات (٢١-١) نتبع الاتى :

$$\text{Var}(\hat{Y}_0 - Y_0) = \text{Var}(\hat{Y}_0) + \text{Var}(Y_0)$$

لأن  $\hat{Y}_0$  و  $Y_0$  مستقلين .

و على ذلك :

$$\text{Var}(\hat{Y}_0 - Y_0) = \sigma^2 \left[ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] + \sigma^2$$

$$= \sigma^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right] .$$

وعلى ذلك الإحصاء:

$$T = \frac{\hat{Y}_0 - Y_0}{\sqrt{S^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right]}} ,$$

يتبع توزيع  $t$  بدرجات حرية  $n-2$  .

يمكن الحصول على  $100\%(1-\alpha)$  فترة ثقة لقيمة مفردة  $y_0$  من الصيغة التالية:

$$\hat{y}_0 - t_{\alpha/2}(n-2) \sqrt{s^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right]}$$

$$< y_0 < \hat{y}_0 + t_{\alpha/2}(n-2) \sqrt{s^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right]} .$$

للبيانات في جدول (١٠-١) و الخاصة بالمثال (٧-١) أوجد 95% فترة ثقة لـ  
 $x = 4$  عند  $y_0$

الحل

$$SXX = 46.91667, \quad n = 12, \quad s^2 = 33.6881, \quad \bar{x} = 4.58333 \quad (١)$$

و على ذلك 95% فترة ثقة لـ  $y_4$  هي :

$$191.84 - 2.228 \sqrt{33.6881 \left[ 1 + \frac{1}{12} + \frac{(4 - 4.58333)^2}{46.91667} \right]} < y_4 <$$

$$191.84 + 2.228 \sqrt{33.6881 \left[ 1 + \frac{1}{12} + \frac{(4 - 4.58333)^2}{46.91667} \right]} .$$

أي أن :

$$191.84 - 2.228(6.061) < y_4 < 191.84 + 2.228(6.061) .$$

والتي تختصر إلى :

$$178.3 < y_4 < 205.3 .$$

وبصورة عامة إذا كان هناك  $m$  من المشاهدات الجديدة فإنه يمكن الحصول  
 علي فترة ثقة تكويه لمتوسط هذه المشاهدات الجديدة عند  $x = x_0$  ويرمز له  
 بالرمز  $\bar{y}$ . التوزيع العيني للإحصاء  $(\hat{Y}_0 - \bar{Y}_0)$  يقع توزيعا طبيعيا بمتوسط :

$$E(\hat{Y}_0 - \bar{Y}_0) = 0 .$$

وتباين :

$$\text{Var}(\hat{Y}_0 - \bar{Y}_0) = \text{Var}(\hat{Y}_0) + \text{Var}(\bar{Y}_0)$$

$$= \sigma^2 \left[ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right] + \frac{\sigma^2}{m}$$

$$= \sigma^2 \left[ \frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right]$$

الآن توزيع الإحصاء :

$$T = \frac{\hat{Y}_0 - \bar{Y}_0}{\sqrt{S^2 \left[ \frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SXX} \right]}}$$

يتبع توزيع  $t$  بدرجات حرة  $n-2$ .

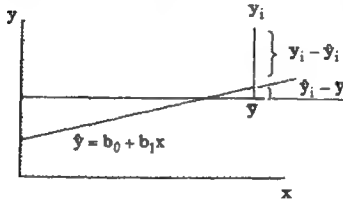
### (١٢-١) أسلوب تحليل الانحدار

#### Analysis of variance approach

تناولنا في البند (١٠-١) استخدام الإحصاء  $T$  لاختبار فرض العدم  $H_0: \beta_1 = 0$  ضد الفرض البديل  $H_0: \beta_1 \neq 0$  وذلك لاختبار جودة خط الانحدار المقدر. غالباً يجرى اختبار الفرض السابق بأسلوب تحليل الانحدار حيث يجزأ الاختلاف الكلي في المتغير التابع إلى مكونات ذات معنى. بفرض لدينا عينة عشوائية من  $n$  نقاط البيانات في الشكل العادي  $(x_i, y_i)$  ولأنه تم تقدير خط الانحدار. في أسلوب تحليل الانحدار سوف نبدأ بالمعادلة التالية:

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i) \quad (٢٢-١)$$

والموضحة بيانياً في الشكل (٢٥-١)



شكل (٢٥-١)

بتربيع طرفي (٢٢-١) والجمع على كل قيم المشاهدات التي عددها  $n$  نحصل على :

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 + 2 \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) \quad (٢٣-١)$$

الحد الثالث على الجانب الأيمن من (٢٣-١) يمكن إعادة كتابته على الشكل التالي:

$$\begin{aligned} 2 \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) &= 2 \sum_{i=1}^n \hat{y}_i (y_i - \hat{y}_i) - 2\bar{y} \sum_{i=1}^n (y_i - \hat{y}_i) \\ &= 2 \sum_{i=1}^n \hat{y}_i e_i - 2\bar{y} \sum_{i=1}^n e_i = 0 . \end{aligned}$$

وذلك لأن مجموع البواقي دائماً تساوي صفر [الخاصية (١) من البند (٧-١)] ومجموع البواقي المرجح بالقيم المقدرة  $\hat{y}_i$  أيضاً يساوي صفر [الخاصية (٤) من البند (٧-١)] وعلى ذلك :

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (٢٤-١)$$

الجانب الأيسر من (٢٤-١) يسمى مجموع المربعات المصحح للملاحظات  $y_i$  (corrected sum of squares) ،  $SYX$  ، والذي يقيس الاختلاف الكلي في المشاهدات  $y_i$ . المكونين على الجانب الأيمن من (٢٤-١) يقسمان على التوالي كمية الاختلاف في المشاهدات  $y_i$  الناتجة من خط الانحدار والاختلاف الباقي والذي لم يفسر بخط الانحدار حيث  $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$  يرمز لمجموع مربعات

الانحدار the regression sum squares و أما  $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$  فيرمز لمجموع مربعات البواقي (١٦-١). وعلى ذلك (٢٣-١) يمكن كتابتها على الشكل التالي:

$$SYX = SSR + SSE . \quad (٢٥-١)$$

وبمقارنة (٢٥-١) بالمعادلة (١٧-١) فإن مجموع المربعات للانحدار يمكن حسابه كالتالي :

$$SSR = b_1 SXY .$$

أن عملية تجزئة درجات الحرية تتم كالتالي . المجموع  $SYX$  له  $n-1$  درجات حرية وذلك لأن درجة حرية واحدة فقدت نتيجة للتقيد  $\sum_{i=1}^n (\hat{y}_i - \bar{y}) = 0$  على الانحرافات  $y_i - \bar{y}$ . المجموع  $SSR$  له درجة حرية واحدة وذلك لأن  $SSR$  يقدر

كاملاً بمعلمة واحدة وهي  $b_1$  . وفي النهاية فإننا نعلم سابقاً أن  $SSE$  لها  $n-2$  درجات حرية وذلك لوجود قيدين على الانحرافات  $y_i - \hat{y}_i$  في عملية تقدير  $b_0, b_1$  . ولأن درجات الحرية لها خاصية التجميع فإن  $n-1=1+(n-2)$  . لاختبار الفرض  $H_0: \beta_1 = 0$  ضد الفرض البديل  $H_1: \beta_1 \neq 0$  فإننا نعلم أنه تحت فرض العدم يمكن إثبات أن  $SSE/\sigma^2$  و  $SSR/\sigma^2$  متغيرين عشوائيين يتبعان مربع كاي بدرجات حرية  $n-2$  و  $1$  على التوالي . أيضاً  $SSY/\sigma^2$  متغير عشوائي يتبع مربع كاي بدرجات حرية  $n-1$  . وعلى ذلك لاختبار الفرض السابق فإننا نستخدم الإحصاء  $F$  والذي يأخذ الشكل التالي :

$$F = \frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{S^2} = \frac{MSR}{MSE} \quad (٢٦-١)$$

حيث  $MSR$  هو متوسط مجموع مربعات الانحدار و  $MSE$  هو متوسط مجموع مربعات البواقي . القيم المتوقعة لكل من  $MSR$  و  $MSE$  سوف تكون :

$$E(MSR) = \sigma^2 + \beta_1^2 SXX, \quad (٢٧-١)$$

$$E(MSE) = \sigma^2 . \quad (٢٨-١)$$

لإثبات (٢٧-١) نتبع الآتي :

$$\text{Var}(B_1) = E(B_1^2) - [E(B_1)]^2 .$$

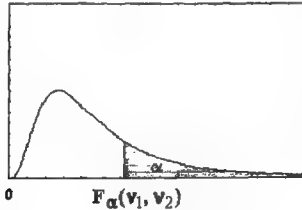
وعلى ذلك :

$$\frac{\sigma^2}{SXX} = E(B_1)^2 - [E(B_1)]^2 .$$

إذن :

$$\begin{aligned} E(MSR) &= E[\Sigma(\hat{Y}_i - \bar{Y})^2] \\ &= E[B_1^2 SXX] \\ &= \left( \frac{\sigma^2}{SXX} + \beta_1^2 \right) SXX \\ &= \sigma^2 + \beta_1^2 SXX . \end{aligned}$$

من (٢٨-١) نجد أن القيمة المتوقعة لمتوسط مربعات البواقي هي  $\sigma^2$  بصرف النظر عن وجود علاقة بين المتغيرين  $Y$  ,  $x$  ، أى أن القيمة المتوقعة تساوى  $\sigma^2$  إذا كانت قيمة المعلمة  $\beta_1$  مساوية للصفر أو تختلف عنه . القيمة المتوقعة لمتوسط مربعات الانحدار تساوى  $\sigma^2$  إذا كانت قيمة  $\beta_1$  مساوية للصفر . أما إذا كانت قيمة المعلمة  $\beta_1$  تختلف عن الصفر فإن القيمة المتوقعة لمتوسط مربعات الانحدار تكون أكبر من ذلك لأن قيمة  $SXX$  و  $\beta_1^2$  موجبة ولذلك كان من الطبيعي أن نقارن بين متوسط مربعات الانحدار  $MSR$  ومتوسط المربعات البواقي  $MSE$  عند اختيار معنوية ميل خط الانحدار . وبما أن  $MSR$  و  $MSE$  متغيرين عشوائيين مستقلين وعندما يكون فرض العدم  $H_0: \beta_1 = 0$  صحيح فإن الإحصاء  $F$  في (٢٦-١) يتبع توزيع  $F$  بدرجات حرية  $n-2$  و  $1$  . إذا كانت قيمة  $F$  المحسوبة كبيرة فإن هذا يعني أن الميل  $\beta_1 \neq 0$  . وعلى ذلك لإختبار الفرض  $H_0: \beta_1 = 0$  فإننا نحسب قيمة للإحصاء  $F$  ونرفض  $H_0$  إذا كانت قيمة  $F$  المحسوبة تزيد عن القيمة الجدولية المستخرجة من جدول توزيع  $F$  من الملحق (٢) عند مستوى معنوية  $\alpha = 0.05$  أو من الملحق (٣) عند مستوى معنوية  $\alpha = 0.05$  . منطقة الرفض موضحة بالمنطقة المظلمة في شكل (٢٦-١) . عادة تلخص الحسابات في جدول تحليل التباين أو اختصاراً جدول ANOVA والموضح في جدول (١٤-١) .



شكل (٢٦-١)

جدول (١٤-١)

Source Of Variance S.O.V	Degrees Of Freedom df	Sum Of Squares SS	Mean Squares MS	F
مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط المربعات	
الانحدار	1	SSR	MSR	$MSR / s^2$
الخطأ	n-2	SSE	$s^2 = \frac{SSE}{n-2}$	
الكلي	n-1	SYY		

مثال (٩-١)

إذا كانت تكاليف صيانة سيارات الثمن تزيد مع عمر السيارة .استخدم البيانات المعطاة في جدول (١٥-١) في:

(أ) تقدير نموذج الانحدار المقدر الخطي

(ب) اختبار فرض العدم  $H_0: \beta_1 = 0$  ضد الفرض البديل  $H_1: \beta_1 \neq 0$  عند مستوى معنوية  $\alpha = 0.05$ .

جدول (١٥-١)

x	4.5	4.5	4.5	4.0	4.0	5.0	5.0	5.5	5.0	5.0	5.0	6.0	6.0
y	619	1049	1033	495	723	681	890	1522	987	1194	163	182	764

الحل

$$n = 13 \text{ و } \bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{64}{13} = 4.92308,$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{10302}{13} = 792.462 ,$$

$$SXY = \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} = 50648.5 - \frac{(64)(10302)}{13} = -69.0385,$$

$$SXX = \sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} = 320 - \frac{(64)^2}{13} = 4.92308,$$

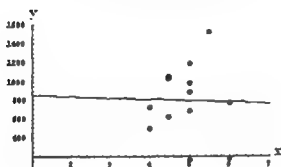
$$b_1 = \frac{SXY}{SXX} = \frac{-69.0385}{4.92308} = -14.0234,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 792.462 - (-14.0234)(4.92308) = 861.5.$$

معادلة الانحدار المقدرة هي :

$$\hat{y} = 861.5 - 14.0234x.$$

والممثلة بيانيا في شكل (٢٧-١) مع شكل الانتشار.



شكل (٢٧-١)

الآن نحسب :

$$SYY = \sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} = 9933940 - \frac{(10302)^2}{13} = 1.77001 \times 10^6$$

$$SSR = b_1 SXY = \frac{(SXY)^2}{SXX} = \frac{(-69.0385)^2}{4.92308} = 968.157$$

ويطرح SSR من SYY نحصل على:

$$SSE = SYY - SSR = 1.77001 \times 10^6 - 968.157 = 1.76904 \times 10^6,$$

$$MSR = \frac{SSR}{1} = \frac{968.157}{1} = 968.157,$$

$$MSE = \frac{SSE}{n-2} = \frac{1769040}{11} = 160821.8.$$



جدول تحليل التباين معطى في جدول (١٦-١)

جدول (١٦-١)

s.o.v	df	SS	MS	F
regression	1	968.157	968.157	8.88682087
residual	11	$1.76904 \times 10^6$	160822	---
Total	12	$1.77801 \times 10^6$	---	---

وبما أن قيمة F المحسوبة من جدول (١٦-١) أقل من قيمة F الجدولية وهى :  $F_{0.05}(1,11) = 4.84$  فإننا نقبل فرض العدم  $H_0: \beta_1 = 0$  . و يجب التنويه هنا أنه عندما تكون قيمة F المحسوبة أقل من الواحد الصحيح فإننا نقبل فرض العدم بدون النظر إلى قيمة F الجدولية .

العلاقة بين F , t

في البند (١٠-١) استخدم الإحصاء :

$$T = \frac{B_1 - \beta_1}{\sqrt{S^2 / SXX}}$$

وذلك لاختبار فرض العدم :

$$H_0: \beta_1 = \beta_1^*$$

ضد الفرض البديل :

$$H_1: \beta_1 \neq \beta_1^* .$$

حيث T تتبع توزيع t بدرجات حرية  $v = n - 2$  . نرفض  $H_0$  عندما  $|t| > t_{\alpha/2}(v)$  وذلك عند مستوى معنوية  $\alpha$  ودرجات حرية  $v$  . في الحالة الخاصة وعند اختبار فرض العدم :

$$H_0: \beta_1 = 0 ,$$

ضد الفرض البديل :

$$H_1: \beta_1 \neq 0 ,$$

فإن القيمة المحسوبة للإحصاء t تصبح :

$$t = \frac{b_1}{\sqrt{s^2 / SXX}}$$

الاختبار في هذه الحالة يكافئ الاختبار الذي نحصل عليه من جدول تحليل التباين المعطى في جدول (١٤-١) وذلك لفرض بديل ذي جانبيين . ويمكن اثبات ذلك كالتالي :

$$t^2 = \frac{b_1^2 SXX}{s^2} = \frac{b_1 SXY}{s^2} = \frac{SSR}{s^2} = \frac{MSR}{MSE} , \quad (٢٩-١)$$

حيث  $t^2$  في (٢٩-١) هي نفسها قيمة F المحسوبة من (٢٦-١) . العلاقات الأساسية بين توزيع t بدرجات حرية  $v$  وتوزيع F بدرجات حرية  $v$  و 1 حيث  $v = n - 2$  هي :

$$t_{\alpha/2}^2(v) = F_{\alpha}(1, v).$$

في الحقيقة ، إن اختبار t يسمح باختبار من جانب واحد بينما اختبار F مفيد في اختبار ذو جانبيين .

#### شكل آخر لجدول تحليل التباين

الشكل الأكثر شيوعاً لجدول تحليل التباين (والذي لن نحتاج له هنا ولكن يكون مفيد في أغراض المقارنة في الفصول القادمة ) يمكن الحصول عليه وذلك بتجزئة مجموع المربعات الغير مصحح  $\sum_{i=1}^n y_i^2$  وذلك من الخطوات التالية :

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

وعلى ذلك :

$$\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

أي أن :

$$\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 = SSR(b_0) + SSR + SSE$$

حيث SSR يرمز له في بعض الأحيان بالرمز  $SSR(b_1 | b_0)$  . وعلى ذلك جدول تحليل التباين يمكن صياغته كما هو موضح في جدول (١٧-١) بالاعتماد

على تجزئة  $\sum_{i=1}^n y_i^2$  .

جدول (١٧-١)

S.O.V	df	SS	MS	F
المحدد $\beta_0$	1	$SSR(\beta_0)$		$F = \frac{MSR}{MSE}$
المحدد $\beta_1   \beta_0$	1	$SSR(\beta_1   \beta_0)$	$MSR = \frac{SSR(\beta_1   \beta_0)}{1}$	
الخطأ	n-2	$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$MSE = \frac{SSE}{n-2}$	
الكل	n			

للبيانات الخاصة بالمثال (٩-١) والمعطاة في جدول (١٥-١) فإننا نحصل على جدول تحليل التباين والمعطى في جدول (١٨-١)

جدول (١٨-١)

S.O.V	df	SS	MS
المحدد $\beta_0$	1	8163938.769	968.157
المحدد $\beta_1   \beta_0$	1	968.157	968.157
الخطأ	11	1769040	160821.8
الكل	13	9933940	

#### (١٣-١) معامل التحديد

#### Coefficient of determination

علمنا في البند (١٢-١) أن :

$$SYY = SSE + SSR \quad (٣٠-١)$$

ويقسمة طرفي المعادلة (٣٠-١) على SYY نحصل على :

$$1 = \frac{SSE}{SYY} + \frac{SSR}{SYY}$$

أي ان:

$$R^2 = \frac{SSR}{SYY} = 1 - \frac{SSE}{SYY} \quad (٢١-١)$$

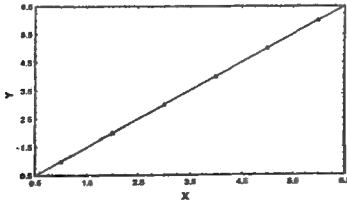
هو معامل التحديد والذي من السهل حسابه من جدول تحليل التباين ويعتبر من أكثر المقاييس الوصفية شيوعا لوصف العلاقة الخطية بين  $x$  و  $y$  وخصوصا في الانحدار المتعدد. بما ان  $0 \leq SSE \leq SYY$  فإن هذا يعني ان  $0 \leq R^2 \leq 1$  ، وقمة  $R^2$  القريبة من 1 تعني ان معظم التغير في  $y$  يفسر من نموذج الانحدار. يمثل معامل التحديد نسبة مساهمة معادلة الانحدار المقدرة في تفسير أو شرح الانحرافات الكلية في قيم  $y$  حول الوسط الحسابي  $\bar{y}$  ، او النسبة بين الانحرافات الموضحة إلى الانحرافات الكلية. فمثلا لو كانت  $R^2 = 0.9$  فهذا يعني ان 90% من الانحرافات الكلية في قيم  $y$  تم شرحها او تفسيرها بواسطة قيم المتغير المستقل  $x$  أو ان معادلة الانحدار المقدرة تقدر 90% من الانحرافات الكلية في قيم  $y$  وان 10% من الانحرافات الكلية لا تزال غير موضحة اذ من المحتمل ان بعض العوامل لم تؤخذ في الاعتبار في صيغة النموذج المقترح أو أن الصيغة المقترحة هي بالأصل غير ملائمة للتعبير عن العلاقة بين المتغيرات مما يؤدي الى وجود انحرافات غير موضحة. عندما تقع كل قيم  $y_i$  للعينة على معادلة الانحدار المقدرة فإن كل  $e_i = y_i - \hat{y}_i$  سوف تساوي صفر وبالتالي فإن  $SSE=0$  وعلى ذلك :

$$SSR = SYY - SSE = SYY .$$

وعلى ذلك :

$$R^2 = \frac{SSR}{SYY} = \frac{SYY}{SYY} = 1 .$$

هذه الحالة موضحة في شكل (٢٨-١) وذلك بالحصول على خط انحدار مضبوط

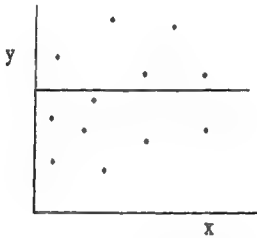


شكل ( ٢٨ - ١ )

ومن ناحية أخرى فإن  $R^2 = 0$  تحدث عندما لا توجد علاقة احصائية بين المتغيرين (حيث  $\hat{y}_i = \bar{y}$  لكل  $i$ ) وعلى ذلك  $SSR=0$  وبالتالي فإن  $SSE$  سوف تساوي  $SYY$  ومنها  $SSR = SYY - SSE = 0$  وعلى ذلك :

$$R^2 = \frac{0}{SYY} = 0$$

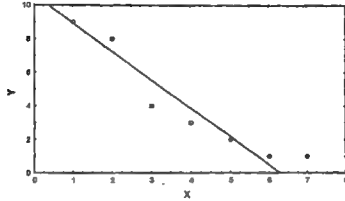
وفي هذه الحالة فإن معادلة الانحدار المقدرة سوف تكون موازية للمحور الافقي ، أي ان  $b_1 = 0$  كما هو موضح في الشكل ( ١ - ٢٩ ).



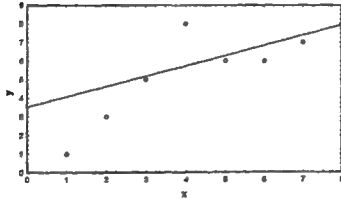
شكل ( ١ - ٢٩ )

عادة في التطبيق  $R^2$  تقع بين 0 و 1 ، عندما تقترب  $R^2$  من 1 فهذا يعني ان هناك درجة من العلاقة الخطية الاحصائية في المشاهدات . كما يتضح من شكل (٣٠-١) حيث  $R^2 = 0.9025$  حيث تقترب المشاهدات بدرجة كبيرة من معادلة الانحدار المقدرة وذلك بالمقارنة مع المشاهدات في شكل (٣١-١) حيث

•  $R^2 = 0.3249$



شكل (٣٠-١)



شكل (٣١-١)

يعتبر  $R^2$  مجرد مقياس وصفي حيث يعتبر الباحثون ان القيم الكبيرة منه دليل على جودة التوفيق لخط الانحدار والقيم الصغيرة من  $R^2$  تعني رداءة في التوفيق . ولكن هذا غير صحيح في كل الاحوال . ويجب استخدام الاحصاء  $R^2$  بشيء من الحذر لانه من الممكن جعل  $R^2$  كبير باضافة حدود كافية الى النموذج . وعلى الرغم من ان  $R^2$  يزيد باضافة متغير مستقل الى النموذج فإن هذا لايعني بالضرورة ان النموذج الجديد اكفىء من النموذج القديم . ايضا قيمة  $R^2$  تعتمد على المدى للمتغير المستقل . عموما  $R^2$  سوف يزيد كلما زاد انتشار قيم  $x$  ويقل كلما قل انتشار قيم  $x$  .

لقد اوضح Hahn (1973) أن القيمة المتوقعة لمعامل التحديد  $R^2$  في حالة الانحدار الخطي البسيط تقريبا تساوي :

$$E(R^2) = \frac{\beta_1^2 SXX}{\beta_1^2 SXX + \sigma^2}.$$

من الواضح ان القيمة المتوقعة لمعامل التحديد سوف تزيد ( تنقص ) عندما  $SXX$  (مقياس الانتشار لقيم  $x$ ) تزيد (تنقص). وعلى ذلك القيمة الكبيرة من  $R^2$  تنتج من ان  $x$  تختلف في مدى كبير. ومن ناحية اخرى فإن قيمة  $R^2$  قد تصبح صغيرة لان مدى  $x$  صغير جدا بدرجة لا تسمح بظهور العلاقة بين  $x$  ,  $y$ . ايضا  $R^2$  لا يقيس النموذج الخطي المناسب فقد تكون قيمة  $R^2$  كبيرة بالرغم من العلاقة بين  $x$  ,  $y$  غير خطية .

يمكن كتابة معامل التحديد على الصيغة التالية :

$$R^2 = \frac{SSR}{SYY} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

أو على الصيغة:

$$= \frac{b_1^2 SXX}{SYY}.$$

أو على الصيغة :

$$R^2 = b_1 \frac{SXY}{SYY}$$

أو على الصيغة:

$$R^2 = \frac{(SXY)^2}{SXX SYY}.$$

$R^2$  و  $F$  العلاقة بين

لتوضيح العلاقة بين  $F$  و  $R^2$  يمكن كتابة الإحصاء  $F$  كما يلي :

$$F = \frac{SSR}{SSE/(n-2)} \quad (٢٢-١)$$

حيث الإحصاء F له درجات حرية  $n-2$  و 1 .

و بقسمة البسط و المقام للمعادلة  $(٢٢-١)$  على  $SSY$  نحصل على :

$$\begin{aligned} F &= \frac{(SSR/SSY)}{(SSE/SSY)/(n-2)} \\ &= \frac{SSR/SSY}{(1 - \frac{SSR}{SSY})/(n-2)} \\ &= \frac{R^2}{(1-R^2)/(n-2)} \end{aligned}$$

للبيانات في المثال  $(١٠-١)$  فإن  $R^2$  تكون :

$$R^2 = 1 - \frac{SSR}{SSY} = 1 - \frac{1.76904 \times 10^6}{1.77001 \times 10^6} = 0.000548 .$$

#### (١٤-١) الانحدار خلال نقطة الاصل

##### Regression through the origin

في كثير من التطبيقات يتطلب حذف  $\beta_0$  من نموذج الانحدار  $(١-١)$ ، أي أن الخط يمر خلال  $x=0$  ,  $y=0$  . هذه الحالة تطبق عند تحليل البيانات في مجال الكيمياء أو في العمليات الصناعية. على سبيل المثال، الاستجابة في عملية كيميائية تساوي صفر عندما تشغل العملية عند درجة حرارة صفر. النموذج في هذه الحالة لا يكون له جزء مقطوع من المحور الرأسي  $y$  ويأخذ الشكل التالي :

$$Y_i = \beta x_i + \epsilon_i \quad (٢٣-١)$$

نفرض أن لدينا  $n$  من أزواج المشاهدات  $(x_i, y_i); i=1,2,\dots,n$  وبما أن  $\beta_0 = 0$  فإن البواقي تأخذ الشكل التالي:  $e_i = y_i - \hat{y}_i = y_i - b_1 x_i$  وعلي ذلك  $SSE$  تأخذ الشكل التالي:

$$SSE = \sum_{i=1}^n (y_i - b_1 x_i)^2 ,$$

المعادلة الطبيعية الوحيدة سوف تكون:



$$b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i .$$

وعلى ذلك تقدير المربعات الصغرى للميل هو:

$$b_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} .$$

المقدر  $B_1$  يكون غير متحيز للمعلمة  $B_1$  ، وتبين  $B_1$  هو:

$$\text{Var}(B_1) = 0 + \frac{\sigma^2 \sum_{i=1}^n x_i^2}{\left( \sum_{i=1}^n x_i^2 \right)^2} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} .$$

ونموذج الانحدار المقدر يأخذ الشكل التالي:

$$\hat{y} = b_1 x .$$

التقدير للتباين  $\sigma^2$  سيكون:

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1} = \frac{\sum_{i=1}^n y_i^2 - b_1 \sum_{i=1}^n x_i y_i}{n-1} .$$

بدرجات حرية  $n-1$  .

مثال (١٠-١)

قيما يلي عدد لوحات الطباعة لمخطوطه (x) والتكلفة الكلية بالدولار لتصحيح الأخطاء المطبعية (y) وذلك لعينه عشوائية من الطلبات الحديثة التي تمهنتها شركه متخصصة في مخطوطات تقنيه. وبما أن Y ينطوي على متغير تكاليف فقد رغب باحث في تحديد ما إذا كان نموذج الانحدار عبر نقطه الأصل (١-٢) ملائما لدراسة العلاقة بين المتغيرين. والبيانات معطاة في جدول (١-١٩) .

جدول (١٩-١)

X	y	x <sup>2</sup>	xy
6	107	36	642
4	75	16	300
10	177	100	1770
18	324	324	5832
25	457	625	11425
30	540	900	16200
25	446	625	11150
14	250	196	3500
10	178	100	1780
10	191	100	1910
12	213	144	2556
7	128	49	896
171	3086	3215	57961

الحل

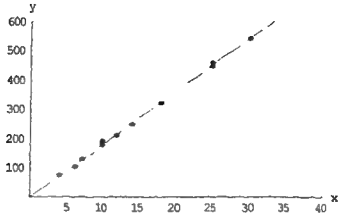
من البيانات في جدول (١٩-١) نحصل على:

$$b_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{57961}{3215} = 18.0283.$$

ومعادلة الاتحاد المقدرة سوف تكون

$$\hat{y} = 18.0283x,$$

والممثلة بيانيا في شكل (٣٢-١) مع شكل الانتشار.



شكل (٣٢-١)

يتضح من شكل (٣٢-١) أن شكل الانتشار يؤكد أن معادله الانحدار المقدرة تمر بنقطة الأصل.

#### معامل التحديد

للمودج (١-١) فإن معامل التحديد يأخذ الشكل التالي :

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} .$$

في حالة النمودج (٣٣-١) فإن المعادلة (٢٤-١) تصبح:

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum (y_i - \hat{y}_i)^2 .$$

و على ذلك فإن معامل التحديد  $R^2$  يصبح :

$$R_0^2 = \frac{\sum_{i=1}^n \hat{y}_i^2}{\sum_{i=1}^n y_i^2} . \quad (٣٤-١)$$

الإحصاء  $R_0^2$  يوضح نسبة الاختلاف (التغير) حول نقطه الأصل (0,0) والنتج من الانحدار. عادة فإن  $R_0^2$  تكون أكبر من  $R^2$  فيما عدا إذا كان متوسط

مجموع مربعات البواقي للنموذج الذي فيه الجزء المقطوع (١-١) أقل من متوسط مجموع مربعات الخطأ للنموذج الذي لا يحتوى على الجزء المقطوع. وهذا يحدث لأن  $R_0^2$  تحسب باستخدام مجموع المربعات الغير مصحح .

للمثال (١٠-١) يتم حساب معامل التحديد من البيانات فى جدول (٢٠-١) كالآتى :

جدول (٢٠-١)

y	y <sup>2</sup>	$\hat{y}$	$\hat{y}^2$
107	11449	108.17	11700.7
75	5625	72.1132	5200.32
177	31329	180.283	32502.
324	104976	324.509	105306.
457	208849	450.708	203137.
540	291600	540.849	292518.
446	198916	450.708	203137.
250	62500	252.396	63703.9
178	31684	180.283	32502.
191	36481	180.283	32502.
213	45369	216.34	46802.8
128.	16384.	126.198	15926.
<b>3086</b>	<b>1045160</b>	<b>3082.84</b>	<b>1044940</b>

وعلى ذلك معامل التحديد  $R_0$  فى (٢٤-١) يحسب كالآتى:

$$R_0 = \frac{\sum_{i=1}^n \hat{y}_i^2}{\sum_{i=1}^n y_i^2} = \frac{1044940}{1045160} = 0.999786 .$$

عادة يفضل استخدم MSE كأساس للمقارنة بين النموذج (١-١) والنموذج (٣٣-١) كما يتضح من المثال التالي:

#### مثال (١١-١)

بفرض تجربة لدراسة العلاقة بين متغيرين أعطيت معادلة الانحدار المقدرة التالية عند فرض عدم وجود جزء مقطوع :

$$\hat{y} = 0.4026x ,$$

$$R_0^2 = 0.9883 , \text{ MSE} = 0.0893 .$$

ولاختبار فرض العدم  $H_0 : \beta_1 = 0$  فإن  $t$  المحسوبة كانت  $t = 91.13$  والتي تكون معنوية عند  $\alpha = 0.01$ . عند توفيق معادلة الانحدار بجزء مقطوع تم الحصول على معادلة الانحدار المقرة التالية :

$$\hat{y} = -0.0938 + 0.4071x .$$

ولاختبار فرض العدم  $H_0 : \beta_0 = 0$  كانت قيمة  $t = -0.065$  (غير معنوية) والتي تعني عدم وجود جزء مقطوع في النموذج وإذا كانت  $\text{MSE} = 0.0931$  و  $R^2 = 0.9997$  حيث  $\text{MSE}$  في هذا النموذج أكبر من النموذج الأول وعلى ذلك نستنتج أن النموذج الأول أفضل من النموذج الثاني بالرغم من أن  $R^2$  أكبر في الثاني عن الأول.

هناك أسلوب آخر لتعريف معامل التحديد  $R^2$  للنموذج (٣٣-١) . واحدة من هذه الطرق هي :

$$R_0'^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} . \quad (٣٥-١)$$

في بعض الأحيان عندما  $\sum_{i=1}^n (y_i - \hat{y}_i)^2$  تكون كبيرة فإن  $R_0'^2$  قد تكون سالبه.

للمثال (١٠-١) فإن قيمة معامل التحديد  $R_0'^2$  في الصيغة (٣٥-١) يحسب من البيانات المعطاه في جدول (٢١-١) حيث:

$$R_0'^2 = 1 - \frac{223.424}{251546} = 0.999112 .$$

جدول (٢١-١)

y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$	$y - \bar{y}$	$(y - \bar{y})^2$
107	108.17	-1.16983	1.3685	-150.167	22550.
75	72.1132	2.88678	8.3335	-182.167	33184.7
177	180.283	-3.28305	10.7784	-80.1667	6426.69
324	324.509	-0.509487	0.259577	66.8333	4466.69
457	450.708	6.29238	39.594	199.833	39933.4
540	540.849	-0.849145	0.721047	282.833	79994.7
446	450.708	-4.70762	22.1617	188.833	35658.
250	252.396	-2.39627	5.7421	-7.16667	51.3611
178	180.283	-2.28305	5.21231	-79.1667	6267.36
191	180.283	10.717	114.853	-66.1667	4378.03
213	216.34	-3.33966	11.1533	-44.1667	1950.69
128.	126.198	1.80187	3.24672	-129.167	16684.

تحت فرض الاعتدال لحد الخطأ فإن يمكن الحصول على فترات ثقة وفترات تنبؤ واختبارات فروض لنموذج الانحدار الذي لا يحتوي على الجزء المقطوع.  $100\%(1 - \alpha)$  فترة ثقة للمعلمة  $\beta_1$  تأخذ الصيغة التالية :

$$b_1 - t_{\alpha/2}(n-1) \sqrt{\frac{s^2}{\sum_{i=1}^n x_i^2}} < \beta_1 < b_1 + t_{\alpha/2}(n-1) \sqrt{\frac{s^2}{\sum_{i=1}^n x_i^2}} .$$

والآن لإيجاد 95% فترة ثقة للمعلمة  $\beta_1$  وباعتماد على البيانات في

جدول (١٩-١) وجدول (٢١-١) و الخاصة بالمثال (١٠-١) نتبع الآتي :

يتم حساب  $s^2$  من الصيغة التالية :

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1} = \frac{223.424}{11} = 20.311$$

وعلى ذلك 95% فترة ثقة للمعلمة  $\beta_1$  تحسب كالآتي :

$$18.0283 - 2.201 \sqrt{\frac{20.311}{3215}} < \beta_1 < 18.0283 + 2.201 \sqrt{\frac{20.311}{3215}}$$

(حيث  $(t_{0.025}(11) = 2.201$  )

أي أن :

$$18.0283 - 2.201(0.07948) < \beta_1 < 18.0283 + 2.201(0.07948)$$

والتي تختصر إلى :

$$17.8512 < \beta_1 < 18.2054 .$$

100%(1-α) فترة ثقة لـ  $\mu_{Y|x_0}$  ، متوسط الاستجابة ، عندما  $x = x_0$  ستكون:

$$\hat{y}_0 - t_{\alpha/2}(n-1) \sqrt{\frac{x_0^2 s^2}{\sum_{i=1}^n x_i^2}} < \mu_{Y|x_0} < \hat{y}_0 + t_{\alpha/2}(n-1) \sqrt{\frac{x_0^2 s^2}{\sum_{i=1}^n x_i^2}} \quad (٣٦-١)$$

للمثال (١٠-١) وباعتماد علي البيانات في جدول (١٩-١) و جدول (٢١-١) فإن 95% فترة ثقة لـ  $\mu_{Y|x_0}$  عندما  $x_0 = 10$  ستكون:

$$180.283 - 2.201 \sqrt{\frac{(10)^2(20.311)}{3215}} < \mu_{Y|10} < 180.283 + 2.201 \sqrt{\frac{(10)^2(20.311)}{3215}}$$

أي أن :

$$180.283 - 2.201(0.7948) < \mu_{Y|10} < 180.283 + 2.201(0.7948)$$

والتي تختصر إلى :

$$178.512 < \mu_{Y|x_0} < 182.054 .$$

100%(1-α) فترة تنبؤ لمشاهدة مستقبلية  $y_0$  عندما  $x = x_0$  هي :

$$\hat{y}_0 - t_{\alpha/2}(n-1) \sqrt{s^2 \left(1 + \frac{x_0^2}{\sum_{i=1}^n x_i^2}\right)} \leq y_0 \leq \hat{y}_0 + t_{\alpha/2}(n-1) \sqrt{s^2 \left(1 + \frac{x_0^2}{\sum_{i=1}^n x_i^2}\right)} .$$

(٣٧-١)

للمثال (١٠-١) وباعتماد علي البيانات في جدول (١٩-١) و جدول (٢١-١) فإن 95% فترة تنبؤ لمشاهدة مستقبلية عند  $x_0 = 10$  هي:

$$180.283 - 2.201 \sqrt{20.311(1 + \frac{10^2}{3215})} < y_{10} < 180.546 + 2.201 \sqrt{20.311(1 + \frac{10^2}{3215})}$$

أي أن :

$$180.283 - 2.201(4.576) < y_{10} < 180.283 + 2.201(4.576).$$

والتي تختصر إلى :

$$170.086 < y_{10} < 190.48.$$

كلا الفترتين (٣٦-١) و (٣٧-١) تتسع كلما زادت قيمة  $x_0$  وبالإضافة إلى ذلك فإن طول فترة الثقة في (٣٦-١) عند  $x = 0$  هو صفر وذلك لأن النموذج يفترض أن متوسط  $y$  عند  $x = 0$  معروف بالتأكيد أنه يساوي صفر. هذا السلوك يختلف عن الملاحظ في النموذج (١-١) والذي يحسوي على الجزء المقطوع. فترة التنبؤ (٣٧-١) لها طول لا يساوي الصفر عندما  $x_0 = 0$  وذلك لأن الخطأ العشوائي للملاحظة المستقبلية لابد أن يأخذ في الحسبان.

#### (١٥-١) الاستدلال أنيا لمعالم النموذج

##### Simultaneous inference on model parameters

بفرض تقدير  $\beta_0, \beta_1$  بمنطقة ثقة مشتركة بحيث أن  $100\%(1-\alpha)$  ثقة أن كلا التقديرين صحيحين . النموذج سوف يكون :

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

$$= \beta'_0 + \beta_1(x - \bar{x}) + \varepsilon .$$

تقديري المربعات الصغرى للمعلمتين  $\beta'_0, \beta_1$  هما على التوالي  $b'_0 = \bar{y}$  ,  $b_1 = \frac{SXY}{SXX}$  أثبتنا من قبل أن المقدرين  $B'_0, B_1$  غير متحيزين

حيث تبين  $\beta_1$  هو  $Var(B_1) = \frac{\sigma^2}{SXX}$  وتبين  $B'_0$  هو  $Var(B'_0) = \frac{\sigma^2}{n}$  تحت فروض الاعتدال العادية ، فإن المتغيرين :

$$\frac{B'_0 - \beta'_0}{(\sigma^2/n)^{\frac{1}{2}}} , \quad \frac{B_1 - \beta_1}{(\sigma^2/SXX)^{\frac{1}{2}}} ,$$

مستقلين ويتبعان التوزيع الطبيعي القياسي . بما أن مربع المتغير العشوائي الطبيعي القياسي يتبع مربع كاي بدرجة حرية واحدة فإن :



$$\left[ \frac{B'_0 - \beta'_0}{(\sigma^2/n)^{\frac{1}{2}}} \right]^2 = \frac{n(B'_0 - \beta'_0)^2}{\sigma^2} \sim \chi^2_{(1)}$$

و

$$\left[ \frac{B_1 - \beta_1}{(\sigma^2/SXX)^{\frac{1}{2}}} \right]^2 = \frac{SXX(B_1 - \beta_1)^2}{\sigma^2} \sim \chi^2_{(1)}.$$

ولان  $B'_0, B_1$  مستقلين ، فإن المتغيرين السابقين اللذين يتبعان مربع كاي أيضا مستقلين . ومن خاصية الجمع لمربع كاي فإن:

$$\frac{n(B'_0 - \beta'_0)^2}{\sigma^2} + \frac{SXX(B_1 - \beta_1)^2}{\sigma^2} \sim \chi^2_{(2)} .$$

الآن توزيع  $\frac{(n-2)S^2}{\sigma^2}$  هو مربع كاي بدرجات حرية  $(n-2)$ ، ويمكن إثبات أن  $S^2$  مستقل عن  $B'_0, B_1$  . وعلى ذلك النسبة :

$$\begin{aligned} & \frac{\frac{1}{2} \left[ n(B'_0 - \beta'_0)^2 / \sigma^2 + SXX(B_1 - \beta_1)^2 / \sigma^2 \right]}{\left[ (n-2)S^2 / \sigma^2 \right] / (n-2)} \quad (٢٨-١) \\ &= \frac{n(B'_0 - \beta'_0)^2 + SXX(B_1 - \beta_1)^2}{2S^2} \end{aligned}$$

يتبع توزيع  $F$  بدرجات حرية  $n-2$  و  $2$  أي  $F[2, n-2]$  . بوضع :

$$\beta'_0 = \beta_0 + \beta_1 \bar{x} \quad \text{و} \quad B'_0 = B_0 + \beta_1 \bar{x} ,$$

في (٢٨-١) ويتبسط المعادلة، نحصل على :

$$\frac{n(B_0 - \beta_0)^2 + 2 \sum_{i=1}^n x_i (B_0 - \beta_0)(B_1 - \beta_1) + \sum_{i=1}^n x_i^2 (B_1 - \beta_1)^2}{2S^2}$$

وبما أن الجلة الاحتمالية التالية :

$$P\left\{\frac{n(B_0 - \beta_0)^2 + 2 \sum_{i=1}^n x_i (B_0 - \beta_0)(B_1 - \beta_1) + \sum_{i=1}^n x_i^2 (B_1 - \beta_1)^2}{2S^2}\right\} \leq F_{\alpha}[2, n-2] = 1 - \alpha,$$

تكون صحيحة لكل قيم  $\beta_0, \beta_1$  حيث  $F_{\alpha}[2, n-2]$  هي قيمة F الجدولية. وعلى ذلك 100% (1- $\alpha$ ) فترة ثقة مشتركة للمعلمتين  $\beta_0, \beta_1$  تكون على الشكل التالي:

$$n(b_0 - \beta_0)^2 + 2 \sum_{i=1}^n x_i (b_0 - \beta_0)(b_1 - \beta_1) + \sum_{i=1}^n x_i^2 (b_1 - \beta_1)^2 \leq F_{\alpha}[2, n-2]. \quad (٢٩-١)$$

المعادلة (٢٩-١) تعرف قطع والذي يتكرر للمعينة سوف يحتوي على  $\beta_0, \beta_1$  أننا باحتمال 100% (1- $\alpha$ ).

#### مثال (١٢-١)

يعطي جدول (٢٢-١) المشاهدات للمتغير التابع Y والمتغير المستقل x وذلك لعينة عشوائية حجمها n=10 والمطلوب إيجاد فترة ثقة مشتركة للمعلمتين  $\beta_0, \beta_1$ .

#### جدول (٢٢-١)

x	y	$x^2$	xy
0.003	90	$9 \times 10^{-6}$	0.27
0.0082	97	0.00006724	0.7954
0.019	107	0.000361	2.033
0.0278	124	0.00077284	3.4472
0.0331	142	0.00109561	4.7002
0.0445	150	0.00198025	6.675
0.0538	172	0.00289444	9.2536
0.0615	189	0.00378225	11.6235
0.0806	209	0.00649636	16.8454
0.1232	253	0.0151782	31.1696
<b>0.4547</b>	<b>1533</b>	<b>0.0326372</b>	<b>86.8129</b>

**الحل**

بما أن :

$$\bar{y} = \frac{\sum y}{n} = \frac{1533}{10} = 153.3 \quad , \quad \bar{x} = \frac{\sum x}{n} = \frac{0.4547}{10} = 0.04547.$$

فإن :

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$= \frac{86.8129 - \frac{(0.4547)(1533)}{10}}{0.0326 - \frac{(0.4547)^2}{10}}$$

$$= \frac{17.1074}{0.011962} = 1430.14,$$

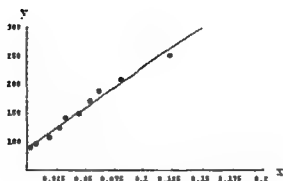
$$b_0 = \bar{y} - b_1 \bar{x} = 153.3 - (1430.14)(0.04547)$$

$$= 88.2714 .$$

معادلة الانحدار المقدرة سوف تكون على الشكل :

$$\hat{y} = 88.2714 + 1430.14 x .$$

والممثلة بيانيا في شكل (٣٣-١) مع شكل الانتشار.



شكل (٣٣-١)

أيضا متوسط مجموع مربعات الخطأ  $s^2$  يحسب من جدول (٢٣-١)

جدول (٢٣-١)

$y$	$\hat{y}$	$(y - \hat{y})$	$(y - \hat{y})^2$
90	92.5619	-2.56186	6.56315
97	99.9986	-2.9986	8.99162
107	115.444	-8.44414	71.3035
124	128.029	-4.02939	16.236
142	135.609	6.39086	40.8431
150	151.913	-1.91276	3.65866
172	165.213	6.78692	46.0622
189	176.225	12.7748	163.196
209	203.541	5.45911	29.8019
253	264.465	-11.4649	131.445
1533			518.101

حيث :

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n-2} = \frac{518.101}{8} = 64.7626,$$

أيضا:

$$b_0 = 88.2714, \quad b_1 = 1430.14$$

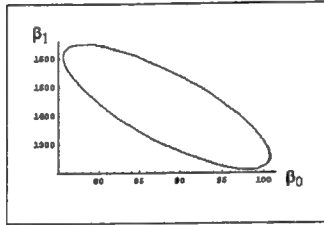
وعلى ذلك فإن منطقة فترة ثقة 95% تعطى على الشكل التالي :

$$\frac{n(b_0 - \beta_0)^2 + 2 \sum_{i=1}^n x_i(b_0 - \beta_0)(b_1 - \beta_1) + \sum_{i=1}^n x_i^2(b_1 - \beta_1)^2}{2s_2^2} \leq F_{\alpha}[2, n-2] .$$

أي أن :

$$\frac{[10(88.2714 - \beta_0)^2 + 2(0.4547)(88.2714 - \beta_0)(1430.14 - \beta_1) + (0.0326)(1430.14 - \beta_1)^2]}{2(64.7626)} = 4.46 ,$$

ومنطقة فترة الثقة موضحة بيانياً في شكل (٣٤-١)



شكل (٣٤-١)

هناك أسلوب عام آخر لإيجاد تقدير بفترة آتيا للمعلمتين  $\beta_0, \beta_1$  في نموذج الانحدار الخطي (١-١). يمكن إيجاد الفترتين كالتالي :

$$b_0 \pm \Delta \sqrt{s^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{SXX} \right]} , \quad (٤٠-١)$$

$$b_1 + \Delta \sqrt{\frac{s^2}{SXX}}, \quad (٤١-١)$$

حيث الثابت  $\Delta$  يختار بحيث أنه لاحتساب خاص فإن كلا الفترتين الناتجتين من العينة نفسها صحيحتان معا . هناك طرق عديدة استخدمت لاختيار  $\Delta$  في (٤٠-١) و (٤١-١) . واحدة من الطرق تسمى طريقة يونفروني . في هذه الطريقة نضع  $\Delta = t_{\alpha/4}(n-2)$  بحيث أن (٤٠-١) و (٤١-١) تصبحان :

$$b_0 \pm t_{\alpha/4}(n-2) \sqrt{s^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{SXX} \right]}, \quad (٤٢-١)$$

$$b_1 \pm t_{\alpha/4}(n-2) \sqrt{\frac{s^2}{SXX}} \quad (٤٣-١)$$

الاحتمال سوف يكون على الأقل  $1-\alpha$  أن كلا الفترتين صحيح . فترات الثقة ليونفروني تشبه فترة ثقة لكل معلمة على حدة والتي تعتمد على توزيع  $t$  فيما عدا أن كل فترة ليونفروني له معامل ثقة  $1 - \frac{\alpha}{2}$  بدلا من  $1-\alpha$  . للتحقق من أن هذه الطريقة تؤدي إلى جمل صحيحة . ليكن  $A_1$  الحادثة أن فترة الثقة للمعلمة  $\beta_0$  غير صحيحة و  $A_2$  الحادثة أن فترة الثقة للمعلمة  $\beta_1$  غير صحيحة . وعلى ذلك  $P(A_1) = P(A_2) = \frac{\alpha}{2}$  . الآن الاحتمال أن واحد أو كلا الجملتين غير صحيح هو :

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) .$$

وعلى ذلك :

$$1 - P(\overline{A_1 \cup A_2}) = 1 - P(A_1) - P(A_2) + P(A_1 \cap A_2) \quad (٤٤-١)$$

وبما أن :

$$1 - P(A_1 \cup A_2) = P(\overline{A_1 \cup A_2}) = P(\overline{A_1} \cap \overline{A_2}) ,$$

(من قانون دي مورجان) فإن الجانب الأيسر من (٤٤-١) هو احتمال أن فترتي الثقة كليهما صحيحتان . وبما أن  $P(A_1 \cap A_2) \geq 0$  فإنه يمكن كتابة (٤٤-١) على الشكل التالي :

$$P(\bar{A}_1 \cap \bar{A}_2) = P(\text{كلا الفترتين صحيحين}) \\ \geq 1 - P(A_1) - P(A_2) \\ \geq 1 - \frac{\alpha}{2} - \frac{\alpha}{2} \geq 1 - \alpha .$$

الصيغة السابقة تسمى متباينة بونفروني. يمكننا بسهولة استخدام متباينة بونفروني للحصول على معامل ثقة مشترك يساوي على الأقل  $1 - \alpha$  وذلك لتقدير  $\beta_0, \beta_1$ . ونقوم بذلك من خلال تقدير  $\beta_0, \beta_1$  بصورة منفصلة وبمعامل ثقة يساوي  $1 - \frac{\alpha}{2}$  لكل منهما وهذا يعطي حد بونفروني  $1 - \alpha = 1 - \frac{\alpha}{2} - \frac{\alpha}{2}$ . لاحظ أن معامل ثقة مساوي  $1 - \frac{\alpha}{2}$  يتطلب استخدام المئين  $100 \left(1 - \frac{\alpha}{4}\right)$  لتوزيع  $t$  وذلك لفترة ثقة ذات جانبيين . وهكذا فإنه إذا قدرت  $\beta_0, \beta_1$  بصورة منفصلة ونقل — 95% فترة ثقة فإن متباينة بونفروني تضمن لنا بمعامل ثقة مشترك يساوي 90% على الأقل أن الفترتين الناتجتين من العينة نفسها صحيحتان معا .

#### مثال (١٣-١)

يعطي الجدول (٢٤-١) السن وضغط الدم لعشرة من الإناث والمطلوب إيجاد 90% فترات ثقة مشتركة للمعلمتين  $\beta_0, \beta_1$  وذلك بالحصول على 95% فترة ثقة لكل معلمة على حده .

جدول (٢٤-١)

x	y	$x^2$	xy
41	124	1681	5084
35	115	1225	4025
62	138	3844	8556
52	149	2704	7748
41	145	1681	5945
58	144	3364	8352
48	145	2304	6960
67	152	4489	10184
66	150	4356	9900
68	150	4624	10200
538	1412	30272	76954

الحل

$$\bar{y} = \frac{\sum y}{n} = \frac{1412}{10} = 141.2 \quad , \quad \bar{x} = \frac{\sum x}{n} = \frac{538}{10} = 53.8 \quad ,$$

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$= \frac{76954 - \frac{(538)(1412)}{10}}{30272 - \frac{(538)^2}{10}}$$

$$= \frac{988.4}{1327.6} = 0.744501 \quad ,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 141.2 - (0.744501)(53.8) = 101.146 \quad .$$

معادلة الانحدار المقدرة سوف تكون على الشكل التالي :

$$\hat{y} = 101.146 + 0.744501x \quad .$$

متوسط مجموع مربعات البواقي  $s^2$  يحسب من جدول (٢٥-١)

جدول (٢٥-١)

y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
124	131.6703	-7.6703	58.8347
115	127.2033	-12.2034	148.9223
138	147.3049	-9.3049	86.5813
149	139.8598	9.1401	83.5414
145	131.6703	13.3296	177.6786
144	144.3269	-0.3269	0.1068
145	136.8818	8.1181	65.9036
152	151.0274	0.9725	0.9459
150	150.2829	-0.2829	0.0800
150	151.7719	-1.7719	3.1396
1412			625.7349



وعلى ذلك :

$$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} \\ = \frac{625.735}{8} = 78.2169 .$$

أيضا :

$$b_0 = 101.146 , \quad \sqrt{\text{Var}(B_0)} = \sqrt{s^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{\text{SXX}} \right]} = 13.3548 ,$$

$$b_1 = 0.744501 , \quad \sqrt{\text{Var}(B_1)} = \sqrt{\frac{s^2}{\text{SXX}}} = 0.242726 ,$$

$$t_{\frac{.05}{4}}(8) = t_{0.025}(8) = 2.306 .$$

$$b_0 - t_{\frac{\alpha}{4}}(n-2) \sqrt{s^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{\text{SXX}} \right]} \leq \beta_0 \leq b_0 + t_{\frac{\alpha}{4}}(n-2) \sqrt{s^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{\text{SXX}} \right]} .$$

أي أن :

$$101.146 - 2.306 \sqrt{78.2169 \left[ \frac{1}{10} + \frac{(53.8)^2}{1327.6} \right]} \leq \beta_0 \leq$$

$$101.146 + 2.306 \sqrt{78.2169 \left[ \frac{1}{10} + \frac{(53.8)^2}{1327.6} \right]} .$$

والتي تختصر إلى :

$$70.3496 \leq \beta_0 \leq 131.942 .$$

و

$$b_1 - t_{\frac{\alpha}{4}}(n-2) \sqrt{\frac{s^2}{\text{SXX}}} \leq \beta_1 \leq b_1 + t_{\frac{\alpha}{4}}(n-2) \sqrt{\frac{s^2}{\text{SXX}}}$$

أي أن :

$$0.744501 - 2.306 \sqrt{\frac{78.2169}{1327.6}} \leq \beta_1 \leq 0.744501 + 2.306 \sqrt{\frac{78.2169}{1327.6}}$$

والتي تختصر إلى :

$$0.184774 \leq \beta_1 \leq 1.30423 .$$

تعتبر طريقة منطقة الثقة المشتركة أكثر كفاءة من طريقة بونفروني وذلك لان المنطقة للقطع دائما أقل من المنطقة في الفضاء المغطى بفترات بونفروني. عادة طريقة بونفروني تكون الأسهل في الحساب. عند إيجاد فترات ثقة باستخدام طريقة بونفروني فإنه عادة لا توجد مستويات معنوية في جداول  $t$  العادية. بعض الآلات الحاسبة الحديثة تعطي قيم  $t_{\alpha}(v)$  عند استدعاء مكتبة الدالة. الجدول في الملحق (٤) يعطي قيم  $t_{\alpha/2m}(v)$  لقيم.

$$\alpha = .1, .05, .01 ; m = 2(1)10 ,$$

$$v = 5(1) 25(5) 50(10) 100 .$$

حيث :

$$P[T \geq t(v)_{\alpha/2m}] = \alpha/2m.$$

هناك طرق أخرى بخلاف طريقة بونفروني وذلك لاختبار  $\Delta$  في  $(-1, 1)$  و  $(-1, 1)$ . في طريقة (Scheffé (1953, 1959) فإن :

$$\Delta = [2F_{\alpha}(2, n-2)]^{\frac{1}{2}} .$$

### (١٦-١) التقدير آليا لمتوسط الاستجابة

#### Simultaneous estimation of mean response

يمكن الحصول على  $m$  من فترات الثقة على متوسط الاستجابة عند فئة من قيم  $x$  الخاصة ولتكن  $x_1, x_2, \dots, x_m$  والتي لها معامل ثقة مشترك على الأقل يساوي  $1 - \alpha$ . الطريقة موضحة بالمثال التالي :

لبيانات المثال (١٣-١) بفرض أننا نرغب في إيجاد 90% فترات ثقة مشتركة على متوسط الاستجابة عند  $x=34$  و  $x=40$ . التقديرات بنقطة ل  $\mu_{Y|x}$  تكون كالتالي:

i	$x_i$	$\hat{y}_{x_i} = 101.146 + 0.744501 x_i$
1	34	126.459
2	40	130.926

معادلة فترة الثقة سوف تكون :

$$\hat{y}_{x_i} - \Delta \sqrt{s^2 \left[ \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SXX} \right]} \leq \mu_{Y|x_i} \leq \hat{y}_{x_i} + \Delta \sqrt{s^2 \left[ \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SXX} \right]}, \quad i = 1, 2.$$

باستخدام بيانات مثال (١٣-١) نجد أن :

$$SXX = 1327.6, \quad \bar{x} = 53.8, \quad s^2 = 78.2169, \quad n = 10.$$

$$\Delta = t_{\frac{\alpha}{2m}}(n-2) = t_{0.025}(8) = 2.306 \quad \text{و } m=2 \quad \text{حيث :}$$

وعلى ذلك :

$$126.459 - (2.306) \sqrt{78.2169 \left[ \frac{1}{10} + \frac{(34 - 53.8)^2}{1327.6} \right]} \leq \mu_{Y|34} \leq$$

$$126.459 + (2.306) \sqrt{78.2169 \left[ \frac{1}{10} + \frac{(34 - 53.8)^2}{1327.6} \right]}.$$

و

$$130.926 - (2.306) \sqrt{78.2169 \left[ \frac{1}{10} + \frac{(40 - 53.8)^2}{1327.6} \right]} \leq \mu_{Y|40} \leq$$

$$130.926 + (2.306) \sqrt{78.2169 \left[ \frac{1}{10} + \frac{(40 - 53.8)^2}{1327.6} \right]}.$$

أي أن 90% فترة ثقة مشتركة لمتوسط الاستجابة تكون :

$$126.459 - (2.306)(5.5605) \leq \mu_{Y|34} \leq 126.459 + (2.306)(5.5605)$$

و

$$130.426 - (2.306)(4.36367) \leq \mu_{Y|40} \leq 130.426 + (2.306)(4.36367)$$

واختصارا فإن :

$$113.636 \leq \mu_{Y|34} \leq 139.281 ,$$

$$120.863 \leq \mu_{Y|40} \leq 140.989$$

(١٧-١) التنبؤ لعدد  $m$  من مشاهدات جديدة

#### Prediction of $m$ new observations

يمكن استخدام إحدى الطرق السابق استخدامها في البند السابق للحصول على فئة من فترات التنبؤ لعدد  $m$  من المشاهدات الجديدة عند مستويات  $x_1, x_2, \dots, x_m$  والتي تعطى ثقة على الأقل تساوي  $1 - \alpha$  . باستخدام بيانات المثال (١٣-١) وبفرض أننا نرغب في الحصول على 90% فترات ثقة مشتركة لمشاهدتين جديدتين عدد  $x_1 = 34, x_2 = 40$  . التقديرين بنقطة لتلك المشاهدتين الجديدتين هما  $\hat{y}_{x_1} = 126.459$  و  $\hat{y}_{x_2} = 130.926$  . لنموذج الحدار خطي و  $m=2$  فإن :

$$\hat{y}_{x_i} - \Delta \sqrt{s^2 \left[ 1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SXX} \right]} \leq y_{x_i} \leq$$

$$\hat{y}_{x_i} + \Delta \sqrt{s^2 \left[ 1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SXX} \right]}, \quad i = 1, 2.$$

باستخدام البيانات للمثال (١٣-١) فإن :

$$126.459 - (2.306) \sqrt{78.2169 \left[ 1 + \frac{1}{10} + \frac{(34 - 53.8)^2}{1327.6} \right]} \leq y_{x_1} \leq$$

$$126.459 + (2.306) \sqrt{78.2169 \left[ 1 + \frac{1}{10} + \frac{(34 - 53.8)^2}{1327.6} \right]}$$

$$130.926 - (2.306) \sqrt{78.2169 \left[ 1 + \frac{1}{10} + \frac{(40 - 53.8)^2}{1327.6} \right]} \leq y_{x_2} \leq$$

$$130.926 + (2.306) \sqrt{78.2169 \left[ 1 + \frac{1}{10} + \frac{(40 - 53.8)^2}{1327.6} \right]}$$

، أن 90% فترة ثقة مشتركة لمشاهدين جديدين تكون :

$$126.459 - (2.306)(10.4468) \leq y_{x_1} \leq 126.459 + (10.4468)(10.4168)$$

$$130.426 - (2.306)(9.86198) \leq y_{x_2} \leq 130.426 + (2.306)(9.86198)$$

تصارأ فإن :

$$102.3687 \leq y_{x_1} \leq 150.5493 ,$$

$$107.684 \leq y_{x_2} \leq 153.17 .$$

### ١٨) التقدير باستخدام الإمكان الأعظم

تستخدم طريقة المربعات الصغرى لتقدير المعالم في نموذج الانحدار الخطي. بصرف النظر عن شكل التوزيع لحدود الخطأ  $\varepsilon$ ، تعتبر طريقة المربعات الأقل مفضلة مقدرات خطية للمعالم  $\beta_0, \beta_1$ . الطرق الإحصائية الأخرى مثل ات الفروض وإيجاد فترات ثقة، تفترض أن حدود الخطأ تتبع توزيعات ع. عندما يكون شكل التوزيع لحدود الخطأ معروف، فإن هناك طريقة بديلة المعالم تسمى طريقة الإمكان الأعظم. لتكن البيانات  $(x_i, y_i)$  حيث  $i = 1, 2, \dots, n$ . تحت فرض أن حدود الخطأ في نموذج الانحدار تتبع التوزيع عي، أي أن  $\varepsilon_i \sim N(0, \sigma^2)$  فإن  $Y_i$  في نموذج الانحدار الخطي (١-١) ع. متغيرات عشوائية تتبع توزيعات طبيعية وتمثل متغيرات عشوائية مستقلة بمتوسط  $\beta_0 + \beta_1 x$  وتباين  $\sigma^2$ . دالة الإمكان يمكن الحصول عليها من دالة للتوزيع المشتركة للمشاهدات. تعتبر أن دالة التوزيع المشتركة إذا علمت المشاهدات لها معالم ثابتة ومجهولة  $\beta_0, \beta_1, \sigma^2$ . لنموذج الانحدار الخطي

البسيط بحدود خطأ تتبع توزيعات طبيعية فإن دالة الإمكان سوف تكون على الشكل التالي :

$$L(y_i, x_i, \beta_0, \beta_1, \sigma^2) = (2\pi\sigma^2)^{\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \right].$$

تقديرات الإمكان الأعظم تمثل قيم المعالم ، أي  $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2$  والتي تؤدي إلى تعظيم  $L$  أو بصورة مكافئة  $\ln L$  . وعلى ذلك :

$$\begin{aligned} \ln L(y_i, x_i, \beta_0, \beta_1, \sigma^2) = \\ - \left( \frac{n}{2} \right) \ln 2\pi - \left( \frac{n}{2} \right) \ln \sigma^2 - \left( \frac{1}{2\sigma^2} \right) \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2, \end{aligned}$$

وتقديرات الإمكان الأعظم  $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2$  يمكن الحصول عليها بحل المعادلات التالية :

$$\begin{aligned} \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0, \\ \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i &= 0, \\ -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 &= 0, \quad (٤٥-١) \end{aligned}$$

الحل للمعادلات في ( ٤٥ - ١ ) هو :

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}, \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \\ \hat{\sigma}^2 &= \frac{\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n}. \end{aligned}$$

نلاحظ أن تقديرات الإمكان الأعظم  $\hat{\beta}_0, \hat{\beta}_1$  هي نفسها تقديرات المربعات الصغرى للمعالم  $\beta_0, \beta_1$  على التوالي . أيضا  $\hat{\sigma}^2$  مقدر متحيز للمعلمة  $\sigma^2$  حيث

التحيز سوف يكون صغير عندما  $n$  تكون كبيرة بدرجة كافية . عادة يستخدم التقدير الغير متحيز للمعلمة  $\sigma^2$  حيث :

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-1} = s^2 .$$

### (١٩-١) الارتباط

#### Correlation

في مناقشتنا للتحدار كنا نفترض أن  $X$  متغير تحت التحكم يقاس بأخطاء يمكن إهمالها وأن  $Y$  متغير عشوائي. تشمل كثير من التطبيقات لتحليل الانحدار على حالات يكون فيها كل من  $X, Y$  متغير عشوائي ومستويات  $X$  لا يمكن التحكم فيها . في تلك الحالات فإننا عادة نفترض أن المشاهدات  $(X, Y)$  حيث  $i = 1, 2, \dots, n$  لهما توزيع احتمالي مشترك. فعلى سبيل المثال بفرض أننا نريد إيجاد معادلة انحدار مقدره للعلاقة بين مبيعات المنتجات ودرجة الحرارة العظمى اليومية. من الواضح أننا لا يمكن التحكم في درجة الحرارة العظمى. نختار عينة عشوائية من  $n$  من الأيام ونلاحظ درجة الحرارة العظمى  $x_i$  ومستوى المبيعات  $y_i$  لكل يوم. وعلى ذلك  $(x_i, y_i)$  لهما توزيع مشترك. عندما  $X, Y$  كلاهما متغيرين عشوائين فإن النموذج (١-١) يصبح  $Y = \beta_0 + \beta_1 X + \varepsilon$  . وعلى ذلك سوف يكون لهما دالة كثافة الاحتمال المشتركة ويمكن إيجاد معامل الارتباط correlation coefficient بين  $X, Y$  من الصيغة التالية:

$$\rho = \frac{\text{Cov}(X, Y)}{\{\text{Var}(X)\text{Var}(Y)\}^{1/2}}$$

حيث  $\text{Cov}(X, Y)$  هو تغاير  $X, Y$  . عندما تكون دالة كثافة الاحتمال المشتركة  $f(x, y)$  من النوع المتصل فإن :

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{x - \mu_X\} \{y - \mu_Y\} f(x, y) dx dy ,$$

$$\text{Var}(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x, y) dx dy ,$$

$$\text{Var}(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_Y)^2 f(x, y) dx dy ,$$

حيث :

$$\mu_X = E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dx dy ,$$

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x,y) dx dy .$$

عندما تكون دالة كثافة الاحتمال المشتركة من النوع المنقطع يستبدل التكامل بالمجموع في الصيغة السابقة. يمكن إثبات أن  $-1 \leq \rho \leq 1$  . الكمية  $\rho$  تعتبر مقياس للارتباط بين المتغيرين  $X, Y$  . على سبيل المثال عندما  $\rho = 1$  فإن  $X, Y$  لهما ارتباط تام موجب ، عندما  $\rho = 0$  فإنه يقال أن المتغيرين غير مرتبطين ، أي لا يوجد علاقة خطية بينهما . وهذا لا يعني أن  $X, Y$  مستقلين . عندما  $\rho = -1$  فإن  $X, Y$  يكونان بينهما ارتباط تام سالب . غالباً في تحليل الاتحاد يفترض أن دالة كثافة الاحتمال المشتركة  $f(x,y)$  للمتغيرين  $X, Y$  تتبع التوزيع الطبيعي الثنائي حيث :

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \left( \frac{x-\mu_X}{\sigma_X} \right) \left( \frac{y-\mu_Y}{\sigma_Y} \right) + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\} ,$$

$$-\infty < x < \infty, -\infty < y < \infty$$

حيث  $\mu_Y, \sigma_Y^2$  هما المتوسط والتباين للمتغير  $Y$  و  $\mu_X, \sigma_X^2$  هما المتوسط والتباين للمتغير  $X$  و:

$$\rho = \frac{E(Y-\mu_Y)(X-\mu_X)}{\sigma_X\sigma_Y} = \frac{\sigma_{12}}{\sigma_X\sigma_Y} ,$$

هو معامل الارتباط بين  $X, Y$  . الحد  $\sigma_{12}$  هو التغاير بين  $X, Y$  . التوزيع الشرطي للمتغير  $Y$  إذا علمت قيمة  $x$  هو :

$$f(y|x) = \frac{1}{\sqrt{2\pi}\sigma_{12}} \exp \left[ -\frac{1}{2} \left( \frac{y-\beta_0-\beta_1x}{\sigma_{12}} \right)^2 \right]$$

حيث :

$$\beta_0 = \mu_Y - \mu_X \rho \frac{\sigma_Y}{\sigma_X} ,$$

$$\beta_1 = \frac{\sigma_Y}{\sigma_X} \rho ,$$

$$\sigma_{12}^2 = \sigma_Y^2 (1-\rho^2) .$$

(٤٦-١)



وعلى ذلك ، التوزيع الشرطي للمتغير  $Y$  إذا علم  $x$  طبيعي بمتوسط الشرطي للمتغير  $Y$  إذا علم  $x$  هو نموذج خط مستقيم . وأكثر من ذلك يوجد علاقة بين معامل الارتباط  $\rho$  والميل  $\beta_1$  من (٤٦-١) نجد أنه عندما  $\rho=0$  فإن  $\beta_1=0$  والتي تعني عدم وجود علاقة خطية بين  $X, Y$  ، أي أن المعلومات عن  $x$  لا تساعد في التنبأ عن  $Y$  . يمكن استخدام طريقة الامكان الأعظم لتقدير المعامل  $\beta_0, \beta_1$  حيث مقدرات الامكان الأعظم للمعامل  $\beta_0, \beta_1$  على التوالي :

$$b_0 = \bar{y} - b_1 \bar{x} , \quad (٤٧-١)$$

$$b_1 = \frac{\sum y_i(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{S_{XY}}{S_{XX}} \quad (٤٨-١)$$

التقديرات في (٤٧-١) و (٤٨-١) هي نفسها التي تم الحصول عليه بطريقة المربعات الصغرى في حالة افتراض أن  $x$  متغير تحت التحكم .

عموما فإن نموذج الانحدار عندما  $X, Y$  متغيرين عشوائيين يتبعان التوزيع الطبيعي الثنائي يمكن تحليله بالطرق السابقة التي استخدمناها عندما كان  $x$  متغير تحت التحكم. وذلك يرجع إلى أن المتغير العشوائي  $Y$  إذا علم  $x$  مستقل ويتبع التوزيع الطبيعي بمتوسط  $\beta_0 + \beta_1 x$  وتباين ثابت  $\sigma_{Y^2}$ . هذه النتائج أيضا تتحقق لأي دالة احتمال مشتركة للمتغيرين  $X, Y$  بحيث أن الدالة الشرطية للمتغير  $Y$  إذا علم  $x$  تكون طبيعية .

يمكن إجراء استدلالات عن معامل الارتباط  $\rho$  في هذا النموذج. التقدير للمعلمة  $\rho$  هو معامل الارتباط البسيط  $r$  حيث :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} .$$

ويجب أن نتذكر أن

$$b_1 = \sqrt{\frac{SYY}{SXX}} r \quad (٤٩-١)$$

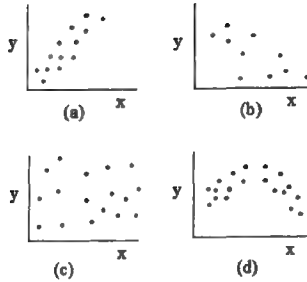
وعلى ذلك الميل  $b_1$  هو معامل الارتباط البسيط  $r$  مضروب في معامل يمثل الجذر التربيعي "لانتشار" قيم  $y$  مقسوما على "انتشار" قيم  $x$ ، وعلى ذلك  $b_1$ ،  $r$  بينهما علاقة بالرغم من أن كل واحد يعطي معلومه مختلفه . فمعامل الارتباط  $r$  هو مقياس للارتباط بين  $X, Y$  بينما  $b_1$  يقيس للتغير المتنبأ به  $y$  عندما تتغير  $x$  بمقدار وحده واحده . عندما تكون  $x$  ثابتة فإن  $r$  لا يكون لها معنى . المعادلة (٤٩-١) تعني أن إشارة معامل الارتباط هي نفس إشارة  $b_1$  .

ايضا يمكن كتابة ، (٤٩-١) كالتالي :

$$r^2 = b_1^2 \frac{SXX}{SYY} \\ = \frac{b_1 SXY}{SYY} = \frac{SSR}{SYY} = R^2.$$

حيث  $R^2$  هو معامل التحديد . أي أن معامل التحديد  $R^2$  هو نفسه مربع معامل الارتباط بين  $X, Y$  . بالرغم من أن الانحدار والارتباط بينهما علاقة قوية فإن الانحدار يعتبر الأداة الأكثر كفاءة في كثير من الحالات . فالارتباط فقط مقياس للارتباط وقليل الاستخدام في التنبؤ.

لاكتشاف العلاقة الخطية يكون من الضروري تمثيل البيانات، إذا كانت نقاط شكل الانتشار تتركز فوق وحول خط انحدار له ميل موجب فهذا يدل على ارتباط موجب قوى بين المتغيرين (ارتباط طردي) كما هو واضح في شكل (١-٣٥)  $a$ . ومن ناحية أخرى إذا كانت نقاط شكل الانتشار تتركز فوق وحول خط انحدار له ميل سالب فهذا يدل على ارتباط قوى سالب بين المتغيرين (ارتباط عكسي) كما هو واضح في شكل (١-٣٥)  $b$ . كلما زاد انتشار نقاط شكل الانتشار حول وفوق خط الانحدار فإن الارتباط يقل عدديا بين المتغيرين . إذا كانت نقاط شكل الانتشار تنتشر بطريقة عشوائية كما في شكل (١-٣٥)  $c$  فهذا يعني أن  $r = 0$  ونستنتج عدم وجود علاقة بين  $X, Y$  . ولما كان معامل الارتباط بين متغيرين يعتبر مقياس للعلاقة الخطية بينهما وعلى ذلك فإن  $r = 0$  تعني قصور في الخطية وليست قصور في الارتباط فعلى سبيل المثال قد تكون هناك علاقة ولكنها علاقة غير خطية . فعلى سبيل المثال إذا وجدت علاقة قوية من الدرجة الثانية بين المتغيرين  $X, Y$  كما هو واضح في شكل (١-٣٥)  $d$  فهذا يعني أن  $r = 0$  .



شكل (١-٣٥)

عادة يتم حساب معامل الارتباط من المعادلة التالية :

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left[ \sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] \left[ \sum y_i^2 - \frac{(\sum y_i)^2}{n} \right]}}$$

$$= \frac{SXY}{\sqrt{SXX.SYY}}$$

مثال (١-١٤)

جدول (١-٢٦) يوضح السن و ضغط الدم لعشرة من الاثناث.

جدول (١-٢٦)

x	41	35	62	52	41	58	48	67	66	68
y	124	115	138	149	145	144	145	152	150	150

المطلوب إيجاد معامل الارتباط الخطي البسيط والبيانات اللازمة معطاه في جدول (٢٧-١).

جدول (٢٧-١)

x	y	x <sup>2</sup>	xy	y <sup>2</sup>
41	124	1681	5084	15376
35	115	1225	4025	13225
52	138	2704	8556	19044
52	149	2704	7748	22201
41	145	1681	5945	21025
58	144	3364	8352	20736
43	145	2304	6960	21025
67	152	4489	10184	23104
55	150	4356	9900	22500
58	150	4624	10200	22500
538	1412	30272	76954	200736

الحل

$$SXY = \sum xy - \frac{\sum x \sum y}{n}$$

$$= 76954 - \frac{(538)(1412)}{10}$$

$$= 988.4 ,$$

$$SXX = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$= 30272 - \frac{(538)^2}{10}$$

$$= 1327.6 ,$$

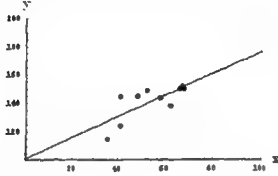
$$SYY = \sum y^2 - \frac{(\sum y)^2}{n}$$

$$= 200736 - \frac{(1412)^2}{10}$$

$$= 1361.6 ,$$

$$r = \frac{SXY}{\sqrt{SXX \cdot SYY}} = \frac{988.4}{\sqrt{(1327.6)(1361.6)}} = 0.735147$$

شكل الانتشار مع معادلة الانحدار المقدرة موضح في شكل (١-٣٦) .



شكل (١-٣٦)

#### اختبارات فروض وفترات ثقه تخصص $\rho$

#### Tests hypotheses and confidence intervals concerning $\rho$

لاختبار فرض العدم  $H_0: \rho = 0$  ضد الفرض البديل  $H_1: \rho \neq 0$  أو الفرض البديل  $H_1: \rho > 0$  أو الفرض البديل  $H_1: \rho < 0$  وبافتراض صحة فرض العدم فإن :

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

هي قيمة لمتغير عشوائي  $T$  له توزيع  $t$  بدرجات حرية  $v = n - 2$ . وعلى ذلك لمستوى معنوية  $\alpha$  وللغرض البديل  $H_1: \rho \neq 0$  (اختبار ذي جانين) فإن منطقة الرفض سوف تكون  $T < -t_{\alpha/2}(v)$  or  $T > t_{\alpha/2}(v)$  حيث  $t_{\alpha/2}(v)$  هي قيمة  $t$  المستخرجة من جدول توزيع  $t$  بدرجات حرية  $v = n - 2$  من الملحق (١) . للبديل  $H_1: \rho > 0$  فإن منطقة الرفض  $T > t_{\alpha}(v)$  وللبديل  $H_1: \rho < 0$  فإن منطقة الرفض  $T < -t_{\alpha}(v)$  .

#### مثال (١-١٥)

لدراسة العلاقة بين تركيز الأوزون  $O_3$  (X) (مقاس PPM) وتركيز الكربون (Y) (مقاس  $g / m^3$ ) تم الحصول على البيانات المعطاة في جدول (١-٧٨) .

جدول (٢٨-١)

x	0.066	0.088	0.120	0.050	0.162	0.186	0.057	0.100
y	4.6	11.6	9.5	6.3	13.8	15.4	2.5	11.8
x	0.112	0.055	0.154	0.074	0.111	0.140	0.071	0.110
y	8.0	7.0	20.6	16.6	9.2	17.9	2.8	13.0

(أ) لوجد معامل الارتباط البسيط .

(ب) وبفرض أن البيانات في مثال (١٥-١) مأخوذة من مجتمع يتبع التوزيع الثنائي الطبيعي. المطلوب اختبار فرض العدم  $H_0: \rho=0$  ضد الفرض البديل  $H_1: \rho > 0$  وذلك عند مستوى معنوية  $\alpha=0.01$  .

الحل

١

$$n = 16, \quad \Sigma x_i = 1.656, \quad \Sigma y_i = 170.6,$$

$$\Sigma x_i^2 = 0.196912, \quad \Sigma x_i y_i = 20.0397,$$

$$\Sigma y_i^2 = 2253.56,$$

$$\begin{aligned} SXY &= \Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n} \\ &= 20.0397 - \frac{(1.656)(170.6)}{16} \\ &= 2.3826, \end{aligned}$$

$$\begin{aligned} SXX &= \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n} \\ &= 0.196912 - \frac{(1.656)^2}{16} = 0.025516, \end{aligned}$$

$$\begin{aligned} SYY &= \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n} = 2253.56 - \frac{(170.6)^2}{16} \\ &= 434.5375. \end{aligned}$$

وعلى ذلك :

$$r = \frac{SXY}{\sqrt{SXX \cdot SYY}} = \frac{2.3826}{\sqrt{(0.025516)(434.5375)}} = 0.716.$$

(ب)

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.716\sqrt{14}}{\sqrt{1-(0.716)^2}} = 3.84.$$

$t_{0.01}(14) = 2.624$  والمستخرجة من جدول توزيع  $t$  في ملحق (١) بدرجات حرية  $v = n - 2 = 16 - 2 = 14$ . منطقة الرفض  $T > 2.624$ . وبما أن  $t$  تقع في منطقة الرفض، نرفض  $H_0$ .

في البند (١٠-١) استخدمنا القيمة  $t = \frac{b_1}{\sqrt{s^2/SXX}}$  لاختبار فرض العدم

$H_0: \beta_1 = 0$ . هنا يمكن إثبات أن  $t = r\sqrt{n-2} / \sqrt{1-r^2} = b_1 / \sqrt{s^2/SXX}$  وهذا يعنى أن الاختبارين متكافئين. وعلى ذلك إذا كان الاهتمام فقط بقياس قوة العلاقة بين متغيرين  $X, Y$  وليس الحصول على معادلة الانحدار الخطي فإن اختبار  $H_0: \rho = 0$  يكون أسهل من اختبار  $t$  لأنه يتطلب كمية قليلة من الحسابات.

استدلالات أخرى تخص  $\rho$ 

الأسلوب المستخدم لاختبار  $H_0: \rho = \rho_0 \neq 0$  عندما لا يكافئ أي طريقة مستخدمة في تحليل الانحدار.

بفرض أن أزواج المشاهدات  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  تمثل عينة عشوائية مأخوذة من مجتمع يتبع التوزيع الثنائي الطبيعي وإذا كانت  $n$  كبيرة وبافتراض صحة فرض العدم فإن:

$$v = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right)$$

هي قيمة لمتغير عشوائي  $V$  تقريباً يتبع للتوزيع الطبيعي بمتوسط:

$$\mu_V = \frac{1}{2} \ln \left( \frac{1+\rho_0}{1-\rho_0} \right) \quad \text{وتباين} \quad \sigma_V^2 = \frac{1}{n-3}, \quad \text{حيث التباين لا يعتمد على } \rho,$$

وعلى ذلك فإن:

$$z = \frac{v - \frac{1}{2} \ln \left( \frac{1+\rho_0}{1-\rho_0} \right)}{1/\sqrt{n-3}}$$

هي قيمة لمتغير عشوائي  $Z$  تقريبا يتبع التوزيع الطبيعي القياسي. الجدول (٢٩-١) يعطى الفروض البديلة ومنطقة الرفض لكل فرض بديل عند مستوى معنوية  $\alpha$ .

جدول ( ٢٩-١ )

الفروض البديلة	منطقة الرفض
$H_1 : \rho \neq \rho_0$	$Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$
$H_1 : \rho > \rho_0$	$Z > z_{\alpha}$
$H_1 : \rho < \rho_0$	$Z < -z_{\alpha}$

مثال (١٦-١)

إذا كان لديك البيانات التالية :

$$n = 20 , \Sigma y_i = 690.30 , \Sigma y_i^2 = 29040.29 ,$$

$$\Sigma x_i y_i = 10818.56 , \Sigma x_i = 285.90 \Sigma x_i^2 = 4409.55,$$

أختبر الفرض  $0.5 < \rho < 0.8$  عند مستوى معنوية  $\alpha = 0.05$

الحل

أي أننا نرغب في اختبار :

$$H_0 : \rho = 0.5 ,$$

$$H_1 : \rho > 0.5$$

$$\alpha = 0.05.$$

وحيث أن  $r = .733$  فإن :

$$v = \frac{1}{2} \ln \left( \frac{1+.733}{1-.733} \right) = .935,$$

$$\mu_v = \frac{1}{2} \ln \left( \frac{1+.5}{1-.5} \right) = .549.$$

وعلى ذلك فإن :



$$z = \frac{v - \frac{1}{2} \ln[(1 + \rho_0)/(1 - \rho_0)]}{1/\sqrt{n-3}}$$

$$= (.935 - .549\sqrt{17}) = 1.59.$$

$z_{0.05} = 1.645$  والمستخرجة من جدول التوزيع الطبيعي القياسي في ملحق (٥).  
منطقة الرفض  $Z > 1.645$  . وبما أن  $z$  تقع في منطقة القبول نقبل  $H_0$  .  
يمكن الحصول على 100%(1- $\alpha$ ) فترة ثقة للمعلمة  $\rho$  من الصيغة التالية :

$$\frac{e^{2c_1} - 1}{e^{2c_1} + 1} \leq \rho \leq \frac{e^{2c_2} - 1}{e^{2c_2} + 1} .$$

$$c_2 = v + \frac{z_{\alpha/2}}{\sqrt{n-3}} , \quad c_1 = v - \frac{z_{\alpha/2}}{\sqrt{n-3}} \quad \text{حيث أن}$$

للبيانات في مثال (١٦-١) فإن :

$$r = 0.733 , \quad v = 0.935 , \quad n = 20,$$

$$c_1 = .935 - 1.96/\sqrt{17} = .460,$$

$$c_2 = .935 + 1.96/\sqrt{17} = 1.410 .$$

يمكن الحصول على 95% فترة ثقة للمعلمة  $\rho$  بالتعويض في الصيغة التالية :

$$\frac{e^{2(.460)} - 1}{e^{2(.460)} + 1} \leq \rho \leq \frac{e^{2(1.410)} - 1}{e^{2(1.410)} + 1} .$$

والتي تختزل إلى :

$$0.43 \leq \rho \leq 0.89 .$$

## الفصل الثاني

مخالفات فروض نموذج الانحدار الخطى البسيط و كيفية اكتشافها و تصحيحها

### Violation in Simple Linear Regression Model: It is Deletions and Correction

مقدمه	(١-٢)
تحليل البواقي	(٢-٢)
خواص البواقي	(١-٢-٢)
رسوم البواقي	(٢-٢-٢)
رسوم بواقي اخرى لإختبار الاعتدال	(٣-٢-٢)
إختبار لنقص الاعتدال	(٤-٢-٢)
إختبار خطية الانحدار	(٣-٢)
تحويلات الى الخط المستقيم	(٤-٢)
اكتشاف وتصحيح عدم ثبات التباين	(٥-٢)
مقدمه	(١-٥-٢)
طرق تحليليه لاكتشاف عدم ثبات التباين	(٢-٥-٢)
تصحيح عدم ثبات التباين	(٣-٥-٢)
اختيار للتحويلات	(٦-٢)
تحويل قيم المتغير التابع	(١-٦-٢)
طرق بيلنيه لتحويل قيم المتغير التابع أو قيمة المتغير المستقل	(٢-٦-٢)
تحويل قيم للمتغير المستقل	(٣-٦-٢)
وجود مشاهدته ولحده أو قليل من المشاهدات المتطرفه	(٧-٢)

## (١-٢) مقدمة

يعتبر شكل الانتشار لازواج المشاهدات  $(x_i, y_i)$  حيث  $i=1,2,\dots,n$  الخطوة الأولى الضرورية في اتخاذ قرار بشأن الشكل الرياضي للعلاقة بين  $x, Y$  في التطبيق وبمجرد توفيق الدالة ذات الشكل المختار يكون من الضروري فحص صلاحية النموذج . في الحقيقة نحتاج الى فحص عدة نماذج انحدار قبل ان تتم عملية الاختيار النهائي . في هذا الفصل سوف نتناول عدة طرق مفيدة لتشخيص ومعالجة الانحرافات (المخالفت) التالية عن نموذج الانحدار الخطي البسيط (١-١).

١. العلاقة بين  $x, y$  ليست خطية.
٢. حدود الخطأ ليست طبيعية.
٣. التباين لحد الخطأ  $\neq$  ليس ثابت .
٤. حدود الخطأ ليست مرتبطه.
٥. التوقع لحد الخطأ  $\neq$  لا يساوي صفر.

وبالرغم من ان دراستنا في هذا الفصل سوف تقتصر على نموذج الانحدار الخطي البسيط الا ان نفس الاسلوب يمكن تعميمه للنماذج التي تحتوي على عدة متغيرات مستقلة . الاداة الاولى لدراسة صلاحية نموذج الانحدار هو تحليل البواقي .

## (٢-٢) تحليل البواقي

### Residuals analysis

يشير تحليل البواقي لفئة من الطرق التشخيصية diagnostic methods لفحص صلاحية نموذج الانحدار وذلك باستخدام البواقي residuals التي تتولناها في الفصل السابق . عندما يكون نموذج الانحدار مناسب للبيانات فيان البواقي  $(y_i - \hat{y}_i)$  حيث  $i=1,2,\dots,n$  سوف تعكس الخواص المفروضة لحدود الخطأ في النموذج .

## (٢-٢-١) خواص البواقي

عندما يفترض في نموذج الانحدار (١-١) أن متغيرات عشوائية تتبع توزيعات طبيعية وتبايناتها ثابتة وعلى ذلك فإن البواقي سوف تظهر بنمط ينسجم مع تلك الخواص. سوف نعتبر البواقي (قبل اختيار العينة) متغيرات عشوائية وعلى ذلك سوف نوجد متوسطاتها وتبايناتها كالتالي :

$$\begin{aligned} E(Y_i - \hat{Y}_i) &= E(Y_i) - E(B_0 + B_1 x_i) \\ &= \beta_0 + \beta_1 x_i - (\beta_0 + \beta_1 x_i) = 0. \end{aligned}$$

وعلى ذلك كل  $(Y_i - \hat{Y}_i)$  له قيمة متوقعة تساوي الصفر . ولأن تركيبة خطية من المتغيرات  $Y_1, Y_2, \dots, Y_n$  فإن  $Y_i - \hat{Y}_i$  أيضا تركيبة خطية من  $Y_1, Y_2, \dots, Y_n$  . وإذا كانت  $Y_1, Y_2, \dots, Y_n$  تتبع توزيعات طبيعية فإن هذا يعني أن كل باقي يتبع توزيعا طبيعيا . تعتبر البواقي غير مستقلة لأنها تتضمن قيم  $\hat{Y}_i$  والتي ترتكز على تقديري العينة  $b_0, b_1$  . وهكذا يرتبط بالبواقي في نموذج الانحدار  $n-2$  درجات حرية .

لنموذج الانحدار الخطي البسيط يمكن إيجاد تباين  $(Y_i - \hat{Y}_i)$  كالتالي :

$$\begin{aligned} \text{Var}(Y_i - \hat{Y}_i) &= \text{Var}(Y_i) + \text{Var}(\hat{Y}_i) - 2\text{Cov}(Y_i, \hat{Y}_i) \\ &= \sigma^2 + \sigma^2 \left[ \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\text{SXX}} \right] - 2\text{Cov}(Y_i, \hat{Y}_i). \end{aligned}$$

الآن يمكن إثبات أن :

$$\begin{aligned} \text{Cov}(Y_i, \hat{Y}_i) &= \text{Cov} \left[ Y_i, \bar{Y} + \frac{\text{SXY}}{\text{SXX}} (x_i - \bar{x}) \right] \\ &= \sigma^2 \left[ \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\text{SXX}} \right]. \end{aligned}$$

وعلى ذلك التباين لـ  $Y_i - \hat{Y}_i$  هو :

$$\text{Var}(Y_i - \hat{Y}_i) = \sigma^2 \left[ 1 - \left( \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\text{SXX}} \right) \right].$$

في بعض الاحيان يكون من المفيد التعامل مع البواقي المعيارية:

$$d_i = \frac{e_i}{\sqrt{\text{MSE}}}, i = 1, 2, \dots, n.$$

حيث :

$$\text{MSE} = \frac{\sum e_i^2}{n - 2}.$$

البواقي المعيارية قبل سحب العينة تعتبر متغيرات عشوائية متوسطة صفر وتباينها تقريبا يساوي واحد.

هناك صيغة اخرى للبواقي وهي بواقي ستوننت والتي تعرف كالتالي :

$$r_i = \frac{e_i}{\sqrt{\text{MSE} \left[ 1 - \left( \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\text{SXX}} \right) \right]}}.$$

$$i = 1, 2, \dots, n.$$

وتعتبر بواقي ستوننت مفيدة في تشخيص الانحرافات عن نموذج الانحدار . غالبا ، في البيانات ذات الحجم الصغير فان بواقي ستوننت تكون اكثر كفاءة من البواقي المعيارية . عندما تكون  $n$  كبيرة سوف يكون هناك اختلاف صغير بين الطريقتين .

مثال (١-٢)

في عملية صناعية اجريت تجربة لدراسة العلاقة بين متغيرين  $x, Y$  والبيانات معطاة في جدول ( ١-٢ ) .

جدول (١-٢)

x	y	x <sup>2</sup>	xy
100.	150.	10000.	15000.
125	140	15625	17500
125	180	15625	22500
150	210	22500	31500
150	190	22500	28500
200	320	40000	64000
200	280	40000	56000
250	400	62500	100000
250	430	62500	107500
300	440	90000	132000
300	390	90000	117000
350	600	122500	210000
400	610	160000	244000
400	670	160000	268000

و المطلوب :

حساب البواقي و البواقي المعيارية و البواقي ستيندنت .

الحل

$$n = 14, \bar{x} = \frac{\sum x}{n} = \frac{3300}{14} = 235.714,$$

$$\bar{y} = \frac{\sum y}{n} = \frac{5010}{14} = 357.857,$$

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{1413500 - \frac{(3300)(5010)}{14}}{913750 - \frac{(3300)^2}{14}}$$

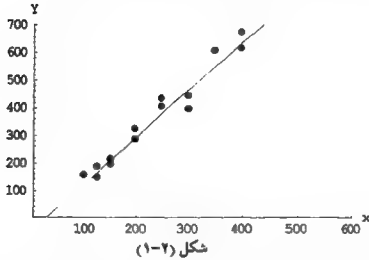
$$= \frac{232571}{135893} = 1.71143,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 357.857 - 1.71143(235.714) = -45.5519.$$

معادلة الانحدار المقدرة هي:

$$\hat{y} = -45.5519 + 1.71143 x.$$

والممثلة بيانيا مع شكل الانتشار في شكل (١-٢).



يعطي جدول (٢-٢) البواقي  $e_i$  والبواقي المعيارية  $d_i$  وبواقي ستونلنت  $r_i$ .

جدول (٢-٢)

$X_i$	$y_i$	$\hat{y}_i$	$e_i$	$d_i$	$r_i$
100	150	125.591	24.4087	0.664208	0.745861
125	140	168.377	-28.3771	-0.772198	-0.843355
125	180	168.377	11.6229	0.316281	0.345426
150	210	211.163	-1.16294	-0.031646	-0.0338405
150	190	211.163	-21.1629	-0.575885	-0.615821
200	320	296.735	23.2654	0.633099	0.660343
200	280	296.735	-16.7346	-0.45538	-0.474977
250	400	382.306	17.6938	0.481484	0.500064
250	430	382.306	47.6938	1.29784	1.34793
300	440	467.878	-27.8778	-0.75861	-0.800463
300	390	467.878	-77.8778	-2.11921	-2.23613
350	600	553.449	46.5506	1.26673	1.38837
400	610	639.021	-29.021	-0.789719	-0.924322
400	670	639.021	30.979	0.842999	0.986682

سوف نتناول في الأجزاء التالية بعض الطرق البيانية والتحليلية لاكتشاف وتصحيح الانحرافات عن نموذج الانحدار الخطي (١-١) وذلك باستخدام البواقي.

### (٢-٢-٢) رسوم البواقي

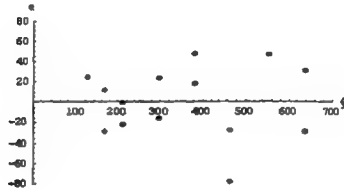
سوف نتناول في هذا الجزء بعض الأنواع من رسوم البواقي (أو رسوم البواقي المعيارية أو رسوم ستيوننت) والتي تستخدم في الكشف عن الانحرافات عن نموذج الانحدار (١-١). كثير من برامج الحاسب الآلي الجاهزة والتي تخص الانحدار تنتج تلك الرسوم حسب الطلب وفي هذه الحالة نحتاج إلى جهد قليل في تشخيص الانحراف عن النموذج.

أ- رسم البواقي مقابل القيم التقديرية:

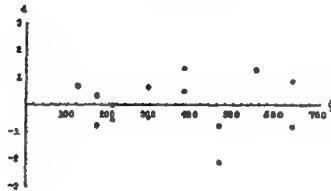
إن رسم البواقي  $e_i$  (أو البواقي المعيارية أو بواقي ستيوننت) مقابل القيم المقدرة  $\hat{y}_i$  يفسر لنا بصورة علمية ما إذا كانت فروض التحليل متوفرة أو لا. إذا



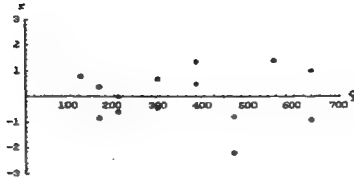
كان النموذج المقدر ملائماً فإن شكل إنتشار البواقي يأخذ الشكل الموضح في شكل (٢-٢) والخاص بالمثال (١-٢) حيث النقاط تنتشر عشوائياً حول الصفر داخل حزام افقى ولا توجد نتوءات أو اتجاه معين كان تصاعدياً أو تنازلياً. نفس الشيء عند استخدام البواقي المعيارية أو بواقي ستويوننت كما هو موضح في شكل (٣-٢) ، (٤-٢) (والخاص بالمثال (١-٢)) على التوالى.



شكل (٢-٢)

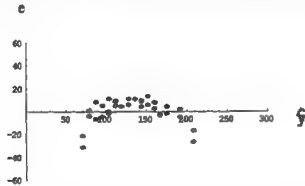


شكل (٣-٢)



شكل (٢-٤)

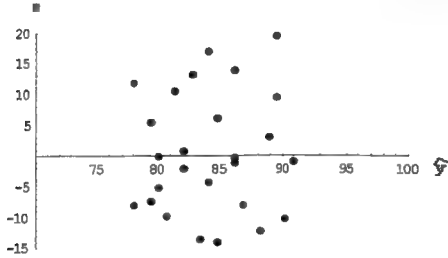
إذا كانت النقاط في رسم البواقي تتوزع على شكل منحنى كما يتضح من شكل (٢-٥) فهذا يدل على عدم الخطية. وهذا يعنى الحاجة الى إضافة متغيرات مستقلة أخرى في النموذج. على سبيل المثال حد التربع قد يكون ضرورياً. التحويلات على المتغير المستقل و (أو) المتغير التابع قد تكون مطلوبة.



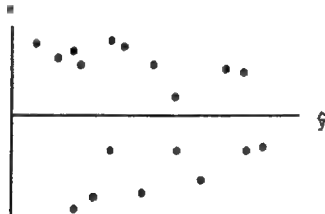
شكل (٢-٥)

الحالة التى يكون فيها فرض التباين غير متحقق موضحة في شكل (٢-٦) حيث يزداد الانتشار الراسى للبواقي مع زيادة  $\hat{y}$  وتسمى هذه الحالة الشكل القمعي المفتوح من الأمام. وهذا يعنى أن توزيعات  $Y_i$  لها تباين يزداد مع زيادة  $\mu_{Y|x_i}$ . أما في شكل (٢-٧) فنجد الانتشار الراسى للبواقي يقل مع زيادة  $\hat{y}$  وتسمى هذه الحالة الشكل القمعي المفتوح من الخلف. وهذا يعنى أن توزيعات  $Y_i$  لها تباين يقل مع زيادة  $\mu_{Y|x_i}$ . وأخيرا شكل (٢-٨) والذي يوضح كلا الشكلين السابقين أى شكل القوس المزدوج وهذا يحدث عندما تكون قيم  $y_i$  نسب تقع بين 0 , 1 حيث تباين نسبة ذى الحدين القريبه من 0.5 يكون اكبر من أخرى قريبه من الصفر أو الواحد

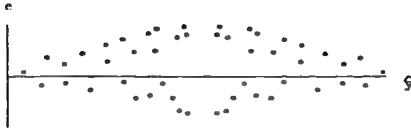
الصحيح، عموماً بفضل استخدام البواقي المعيارية أو بواقي سيتودنت في رسم البواقي.



شكل (٦-٢)



شكل (٧-٢)



شكل (٨-٢)

أيضاً رسم البواقي  $e_i$  مقابل  $\hat{y}$  قد يكشف لنا عن المشاهدات الشاذة (الخوارج) والتي تمثل مجموعة قليلة من المشاهدات في العينة. أن وجود بيانات شاذة في العينة قد يؤدي إلى التوصل إلى نتائج خاطئة. إذا بدأ لنا من شكل الانتشار أن هناك نقطة أو عدة نقاط تبعد بصورة واضحة عن بقية القيم فإن هذه النقطة أو النقاط تمثل بيانات شاذة يستدعي دراستها.

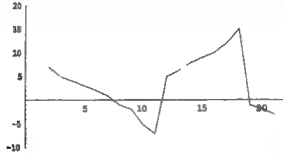
أيضاً رسوم البواقي المعيارية أو بواقي ستوننت تكون مفيدة في اكتشاف الانحراف عن الاعتدال. عندما يكون توزيع الأخطاء طبيعي فإن تقريباً 68% من البواقي المعيارية سوف تقع بين  $+1$ ,  $-1$ ، وتقريباً 95% منهم يقع بين  $+2$ ,  $-2$ ، وما يزيد أو يقل عن ذلك يعتبر أخطاء شاذة (الخوارج outliers).

#### ب- رسم البواقي مقابل متغير مستقل:

عند رسم البواقي  $e_i$  مقابل  $x_i$  وعندما يكون النموذج مناسباً فإن النقاط على الرسم تتبعثر عشوائياً داخل حزام الصفري حول الصفر دون أن تظهر اتجاهات منتظمة لأن تكون موجبه او سالبه. أن رسم البواقي  $e_i$  مقابل قيم  $x_i$  يكافئ رسم البواقي  $e_i$  مقابل القيم التقديرية  $\hat{y}$  وذلك لأن القيم التقديرية  $\hat{y}$  تمثل دوال خطية في القيم  $x_i$  للمتغير المستقل والذي يتأثر فقط هو تدرج محور  $x$  وليس النسب الأساسي للنقاط المرسومة.

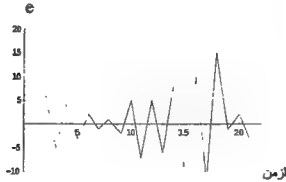
#### ج- رسم البواقي مقابل زمن:

بعض التطبيقات في الاحدار تشتمل علي متغير تابع ومتغيرات مستقلة لها طبيعة ان تكون متتابعه مع الزمن. البيانات في هذه الحالة تسمى السلاسل الزمنية. نماذج الاحدار التي تستخدم السلاسل الزمنية تنتشر في مجال الاقتصاد. إن فرض عدم الارتباط أو الاستقلال للاخطاء لبيانات السلاسل الزمنية يكون غالبا غير منطقي. عادة الاخطاء في السلاسل الزمنية تكون مرتبطة، أي أن  $E(\epsilon_i \epsilon_j) \neq 0$  و  $i \neq j$ . يقال لحدود الخطأ في هذه الحالة انها مرتبطة ذاتيا. في هذه الحالة فإن رسم البواقي  $\epsilon_i$  مقابل الزمن يكشف عن وجود الارتباط الذاتي للبواقي. يتضح من شكل (٩-٢) وجود ارتباط ذاتي موجب حيث تكون هناك عدة نقاط موجبه تليها عدة نقاط سالبه.



شكل (٩-٢)

أما شكل (١٠-٢) فيوضح وجود ارتباط ذاتي سالب حيث نقاط البواقي تتعاقب بالإشارة فالأولى موجبه مثلا والثانيه سالبه والثالثه موجبه والرابعه سالبه وهكذا.



شكل (١٠-٢)

## مثال (٢ - ٢)

يُتوقع أن تقل كتلة عضلات الشخص مع العمر ، ولتقصي هذه العلاقة عند النساء . اختار باحث تغذية أربعة نساء عشوائيا من كل شريحة عمرية من 10 سنوات تبدأ بالعمر 40 وتنتهي بالعمر 79. يعطي جدول (٢ - ٣) النتيجة ،  $x$  العمر و  $y$  قياس كتلة العضلة . بافتراض نموذج الانحدار الخطي البسيط (١ - ١) .

(١) أوجد معادلة الانحدار المقترنة .

(ب) احسب البواقي والبقاوي المعيارية وبواقي ستوننت ومثلها بيانيا. هل تبدو داله الانحدار الخطية توفيقا جيد.

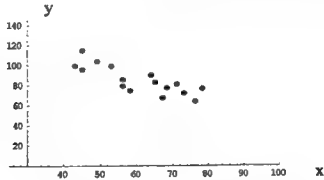
## جدول (٢-٣)

x	71	64	43	67	56	73	68	56	76	65	45	58	45	53	49	78
y	82	91	100	68	87	73	78	80	65	84	116	76	97	100	105	77

## الحل

(١) يتضح من شكل الانتشار (١١-٢) أن الخط المستقيم هو أفضل طريقة لتمثيل هذه البيانات :

أي أننا نفترض النموذج الخطي البسيط .



شكل الانتشار (١١-٢)

بما أن  $\beta_0$  ,  $\beta_1$  مجهولتان فإننا نقدرهما من مشاهدات العينة حيث :

$$n = 16 \quad \sum x_i = 967 \quad \sum x_i^2 = 60409, \\ \sum x_i y_i = 81331 \quad , \quad \bar{x} = 60.4375 \quad , \quad \bar{y} = 86.1875 \quad , \quad \sum y_i = 1379.$$

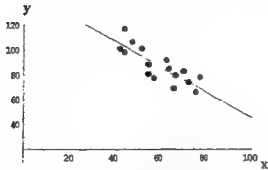
$$b_1 = \frac{SXY}{SXX} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \\ = \frac{81331 - \frac{(967)(1379)}{16}}{60409 - \frac{(967)^2}{16}} \\ = \frac{-2012.3125}{1965.9375} = -1.02359 ,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 86.1875 - (-1.02359)(60.4375) = 148.051 .$$

معادلة الانحدار المقطرة سوف تكون على الشكل :

$$\hat{y} = 148.051 - 1.02359 x.$$

والموضحة بيانياً مع شكل الانتشار في شكل (٢ - ١٢) .



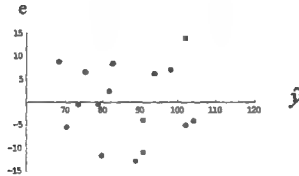
شكل (٢ - ١٢)

(ب) البراقي والبراقي المعيارية وبراقي ستودنت مطاة في جدول (٢-٤)

## جدول (٤-٧)

$y_i$	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$d_i$	$r_i$
82	75.3758	6.6241	43.8795	0.7939	0.8223
91	82.5409	8.4590	71.5553	1.0138	1.0480
100	104.0363	-4.0363	16.2920	-0.4837	-0.4972
68	79.4701	-11.4701	131.5653	-1.3747	-1.4223
87	90.7296	-3.7296	13.9104	-0.4470	-0.4611
73	73.3286	-0.3286	0.1080	-0.0393	-0.0408
78	78.4466	-0.4466	0.1994	-0.0535	-0.0553
80	90.7296	-10.7296	115.1259	-1.2859	-1.3265
65	70.2578	-5.2578	27.6454	-0.6301	-0.6535
84	81.5173	2.4826	6.1634	0.2975	0.3076
116	101.9891	14.0108	196.3036	1.6792	1.7270
76	88.6824	-12.6824	160.8457	-1.5199	-1.5688
97	101.9891	-4.9891	24.8917	-0.5979	-0.6149
100	93.8004	6.1995	38.4344	0.7430	0.7658
105	97.8948	7.1051	50.4838	0.8515	0.8767
77	68.2107	8.7892	77.2515	1.0533	1.0931

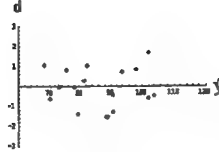
رسم البواقي  $e_i$  مقابل  $\hat{y}_i$  موضح في شكل (١٣-٢).



شكل (١٣-٢)

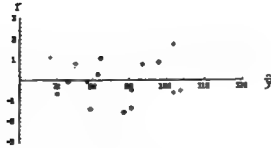
رسم البواقي المعيارية  $d_i$  مقابل  $\hat{y}_i$  موضح في شكل (١٤-٢).





شكل (١٤-٢)

رسم بواقي ستويوننت  $\eta$  مقابل  $\hat{y}$  موضع في شكل (١٥-٢) .



شكل (١٥-٢)

يتضح من شكل الانتشار (١٢-٢) ومن رسوم البواقي أن المعادلة المقتره تبدو توفيقاً جيد.

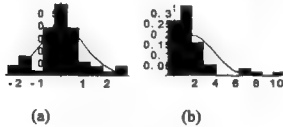
### (٣-٢-٢) رسوم بواقي اخرى لاختبار الاعتدال

بالرغم من أن الانحراف عن الاعتدال لا يؤثر كثيراً على النموذج فإن عدم الاعتدال يؤثر كثيراً في إحصاءات  $F$ ،  $t$ ، وبالتالي على فترات الثقة واختبارات الفروض و التي تعتمد على فرض الاعتدال. أكثر من ذلك فإن الأخطاء التي تأتي من توزيع له ذيل أوسع أو أضيق من الطبيعي يكون توفيق المربعات الصغرى لها حساس للفئات الصغيرة من البيانات . أن توزيعات الأخطاء التي لها ذيل أوسع من

الطبيعي غالبا تنتج من قيم شاذة (الخوارج outliers). في هذا القسم سوف نقدم رسوم بواقي أخرى و ذلك لاختيار ما إذا كانت حدود الخطأ تتبع توزيعات طبيعية عندما يكون هو المطلوب في نموذج الانحدار (١-١).

#### أ- المدرج التكراري

يمكن استخدام المدرج التكراري للبواقي للتحقق من فرض الاعتدال. عندما يكون عدد البواقي صغير جدا فإنه لا يسمح بالتعرف البصري بسهولة على شكل التوزيع الطبيعي. يتضح من شكل (١٦-٢) (a) أن فرض الاعتدال متحقق بينما يوضح شكل (١٦-٢) (b) أن توزيع الأخطاء ملتوي ناحية اليمين.



شكل (١٦-٢)

#### ب- رسم الاحتمال الطبيعي

عموما يستخدم الورق الاحتمالي الطبيعي في تقييم فرض تبعية البيانات للتوزيع الطبيعي حيث توقع المشاهدات المطلوب اختبارها مع القيم المتوقعة الحصول عليها عندما يكون التوزيع طبيعي. فإذا وقعت أزواج القيم الناتجة على خط مستقيم تقريبا فإن هذا يدل على أن البيانات تتبع التوزيع الطبيعي. أما إذا انحرفت النقاط عن خط مستقيم بصورة واضحة فإن فرض الاعتدال يصبح مشكوكا في صحته. كما أن الطريقة التي تحدث بها هذه الانحرافات قد تمدنا ببعض المعلومات عن أسباب عدم التبعية للتوزيع الطبيعي. وبمجرد معرفة هذه الأسباب فإنه من الممكن اتخاذ بعض الإجراءات التصحيحية.

ليكن:

$$Z_{(1)} < Z_{(2)} < \dots < Z_{(n)}$$

الترتيبات الإحصائية و المأخوذة من  $n$  من المتغيرات العشوائية المستقلة والتي لها نفس التوزيع الطبيعي القياسي  $N(0,1)$  فإن القيم المتوسطة للمتغيرات  $Z_{(i)}$  تقريبا تساوي :

$$E(Z_{(i)}) \approx \gamma_{(i)} = \Phi^{-1} \left[ \frac{i-0.5}{n} \right]$$

حيث  $\Phi^{-1}$  هي الدالة العكسية ( $\Phi(z) = a \Rightarrow z = \Phi^{-1}(a)$ ) للتوزيع الطبيعي القياسي حيث:

$$\Phi(z) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt.$$

عندما :

$$U_{(1)} < U_{(2)} < \dots < U_{(n)}$$

تمثل الترتيبات الإحصائية و الناتجة من متغيرات عشوائية مستقلة و لها نفس التوزيع الطبيعي ( $N(\mu, \sigma^2)$ ) فإن:

$$\frac{E(U_{(i)}) - \mu}{\sigma} \approx \gamma_{(i)} \quad (٢-١)$$

و تبعا لذلك فإن :

$$E[U_{(i)}] \approx \mu + \sigma \gamma_{(i)}$$

أن رسم  $U_{(i)}$  مقابل  $\gamma_{(i)}$  سوف يؤدي تقريبا إلى خط مستقيم . ليكن  $e_1, e_2, \dots, e_n$  تمثل البواقي التي عددها  $n$  و بغرض أن  $e_{(1)} < e_{(2)} < \dots < e_{(n)}$  تمثل البواقي

بعد ترتيبها تصاعدياً. أي أن  $e_{(1)}$  أصغر قيمة في البواقي و  $e_{(n)}$  أكبر قيمة وإذا كانت قيم البواقي مختلفة عن بعضها البعض فإنه يوجد عدد  $i$  من البواقي أقل من أو يساوي  $e_{(i)}$ . وهذا الفرض صحيح دائماً من الناحية النظرية إذا كان  $e_i$  متغيراً

عشوائياً مستمراً. وغالباً ما يستخدم المقدار  $p_{(i)} = \frac{(i - \frac{1}{2})}{n}$  كتقريب لنسبة البواقي (في العينة)  $\frac{i}{n}$  التي تقع عند أو على يسار  $e_{(i)}$ . بفرض أن المقدار  $p_{(i)}$  معرف من التوزيع الطبيعي القياسي من العلاقة التالية :

$$\Phi(\gamma(i)) = P(Z \leq \gamma(i)) = \int_{-\infty}^{\gamma(i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = p_{(i)}.$$

وهذا فإن  $p_{(i)}$  هو احتمال الحصول على القيمة أقل من أو يساوي  $\gamma(i)$  وذلك باستخدام التوزيع الطبيعي القياسي. ويرسم أزواج القيم  $(e_{(i)}, \gamma(i))$  وإذا كانت العلاقة بين أزواج القيم خطية تقريباً. فإن هذا يدل على تحقق فرض الاعتدال لحدود الأخطاء. وهناك ورق احتمال طبيعي مصمم بحيث أنه عند رسم  $e_{(i)}$  مقابل  $i = 1, 2, \dots, n, p_{(i)}$  فإن النقاط على الرسم لابد أن تقع قريبه من خط مستقيم وذلك عند تحقق الاعتدال لحدود الأخطاء.

ويمكن إجراء الحسابات اللازمة للحصول على شكل رسم الاحتمال الطبيعي باستخدام الحسبات الآلية. والخطوات اللازمة للحصول على هذا الشكل هي :

١- ترتيب البواقي للحصول على  $e_{(1)}, e_{(2)}, \dots, e_{(n)}$  ثم الحصول على الاحتمالات التجريبية التجميعية لها وهي :

$$(1 - \frac{1}{2}) / n, (2 - \frac{1}{2}) / n, \dots, (n - \frac{1}{2}) / n.$$

٢- حساب قيم  $\gamma(1), \gamma(2), \dots, \gamma(n)$  باستخدام التوزيع الطبيعي القياسي .

٣- رسم أزواج القيم  $(e_{(1)}, \gamma(1)), (e_{(2)}, \gamma(2)), \dots, (e_{(n)}, \gamma(n))$  ثم تحديد ما إذا كانت تقع على خط مستقيم أم لا .

## مثال (٣-٢)

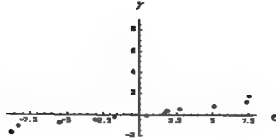
يعطي جدول (٥-٢) القيم المرتبة  $e_{(i)}$  للبواقي الخاصه بالمثال (٧-١) وذلك باستخدام برنامج منفذ على الحاسب الآلي باستخدام برنامج Mathematica مع القيم  $p_{(i)}$  وقيم التوزيع الطبيعي القياسي  $\gamma_{(i)}$  لهذا الاحتمال فمثلا من جدول (٥-٢) نجد أن :

$$P(Z \leq 1.73166) = \int_{-\infty}^{1.73166} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 0.958333.$$

يوضح شكل (١٧-٢) توقييع البواقي المرتبة  $e_{(i)}$  مع قيم التوزيع الطبيعي القياسي  $\gamma_{(i)}$  لنحصل في النهاية على رسم الاحتمال الطبيعي. نلاحظ من شكل (١٧-٢) أن أزواج القيم  $(e_{(i)}, \gamma_{(i)})$  تقع تقريبا على خط مستقيم وبالتالي فإننا نقبل فرضيه تبعيه حدود الخطأ للتوزيع الطبيعي.

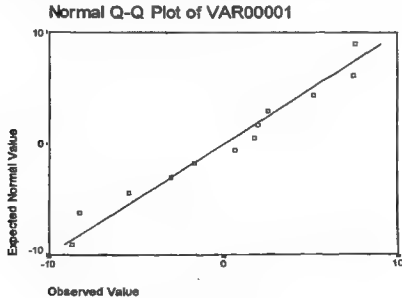
جدول (٥-٢)

$e_i$	$e_{(i)}$	$P_{(i)}$	$\gamma_{(i)}$
5.15808	-8.66963	0.0416667	-1.73166
-8.66963	-8.25577	0.125	-1.15035
-3.01421	-5.42806	0.208333	-0.812218
-8.25577	-3.01421	0.291667	-0.548522
1.91652	-1.66963	0.375	-0.318639
-1.66963	0.571936	0.458333	-0.104633
0.571936	1.74423	0.541667	0.104633
7.50266	1.91652	0.625	0.318639
1.74423	2.74423	0.708333	0.548522
-5.42806	5.15808	0.791667	0.812218
7.39964	7.39964	0.875	1.15035
2.74423	7.50266	0.958333	1.73166

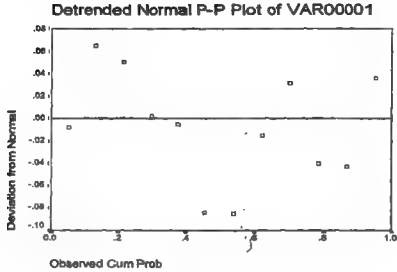


شكل (١٧-٢)

كثير من برامج الحاسب الآلي الإحصائية الجاهزة يمكن استخدامها في رسم البواني علي ورق الاحتمال الطبيعي مثل برامج SPSS و Statistica. يوضح شكل (١٨-٢) رسم احتمال طبيعي خالص بالمثال (٣-٢) منفذ علي الحاسب الآلي باستخدام برنامج SPSS حيث النقاط تقريبا تقع حول خط مستقيم. الانحرافات عن الخط المستقيم في رسم الاحتمال الطبيعي (١٨-٢) موضحة في شكل (١٩-٢).

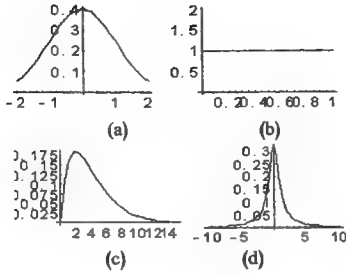


شكل (١٨-٢)

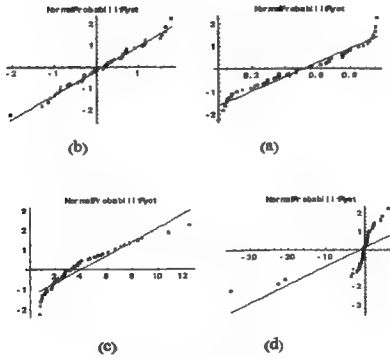


شكل (٢-١٩)

عندما يكون توزيع حدود الخطأ طبيعي كما في شكل (٢-٢٠) 'a' فإننا نحصل على رسم احتمال طبيعي مثالي كما هو موضح في شكل (٢-٢١) b حيث تلتف النقاط حول خط مستقيم. عندما يكون التوزيع ملتوي ناحية اليمين كما في شكل (٢-٢٠) c فإن رسم الاحتمال الطبيعي سوف يكون مقعرا من أسفل downward كما في شكل (٢-٢١) c. أما إذا كان للتوزيع ملتوي ناحية اليسار فإن رسم الاحتمال الطبيعي يكون مقعرا من اعلي upward. وإذا كان التوزيع له احتمال اعلي في الذيلين من التوزيع الطبيعي مثل التوزيع المنتظم كما في شكل (٢-٢٠) b أو التوزيع المقطوع فإن الرسم على الزرق الاحتمالي الطبيعي يكون مقعرا من أسفل ناحية الركن الأيسر السفلي ومقعرا من اعلي ناحية الركن الأيمن العلوي كما في شكل (٢-٢١) a. الحالة العكسية معطاه كما في شكل (٢-٢١) d والتي يمكن ملاحظتها في التوزيعات التي لها احتمال أقل في الذيلين من للتوزيع الطبيعي أو للمدببه والموضحه في شكل (٢-٢١) d



شكل (٢-٢)



شكل (٢-١)



ولأن العينات المأخوذة من توزيع طبيعي لا تقع بالضبط على خط مستقيم ، فإن بعض الخبرة تكون مطلوبة لتفسير رسم الاحتمال الطبيعي. قدم Daniel and wood (1980) عدد كبير من الرسوم على ورق احتمال طبيعي والتي تساعد الباحث في الكشف عن اعدالية توزيعات الأخطاء للعينات من الحجم 8-384 . إن دراسة تلك الرسوم سوف يساعد في تحديد مقدار الانحراف عن الخط المستقيم الذي يمكن قبوله. عادة العينات الصغيرة  $n \leq 16$  تعطي رسم احتمال طبيعي ينحرف بدرجة كبيرة عن الخط المستقيم. للعينات الكبيرة  $n \geq 32$  فإن الرسوم يكون لها سلوك أفضل. عادة نحتاج إلى حوالي 20 نقطة للحصول على رسوم احتمال طبيعي ثابتة بدرجة يسهل تفسيرها. أوضح Andrews (1979) و Gnanadesikan (1977) أن رسوم الاحتمال الطبيعي قد تعطي سلوك طبيعي حتى إذا كانت حدود الخطأ  $\epsilon_j$  لا تتبع التوزيع الطبيعي. وهذه المشكلة تحدث عندما تكون البواقي غير مثالية لعينة عشوائية بسيطة . المخالفة العلة والتي يمكن ملاحظتها في رسم الاحتمال الطبيعي هو وجود واحد أو أكثر من القيم الكبيرة من البواقي في الركن الأيمن الطوي. فسي بعض الاحيان يدل ذلك على وجود مشاهدات شاذة outliers.

#### (٢-٢-٤) اختبار لنقص الاعتدال

في هذا الاختبار يتم ترتيب البواقي المعيارية  $d_{(i)}$  من الأصغر إلى الأكبر (ترتيباً تصاعدياً) حيث  $d_{(1)} < d_{(2)} < \dots < d_{(n)}$  ثم يتم حساب القيم  $z_{(1)} < z_{(2)} < \dots < z_{(n)}$  والمستخرجه من جدول التوزيع الطبيعي القياسي في ملحق (٥) والتي السماح قبلها تسالوى  $p_{(i)} = [(i - 0.75) / (n + 0.25)]$  حيث  $i = 1, 2, \dots, n$  ويمكن استخدام الحاسب الآلي باستخدام برنامج Mathematica في حساب قيم  $z_{(i)}$ . لآزواج القيم :

$$(d_{(1)}, z_{(1)}) , (d_{(2)}, z_{(2)}) , \dots , (d_{(n)}, z_{(n)})$$

يتم حساب معامل الارتباط البسيط  $r$  . بفرض أن فرض النجم:

$H_0$ : توزيع حدود الخطأ في توزيع الانحدار الخطى البسيط (١-١) طبيعي  
ضد الفرض البديل:

$H_1$ : توزيع حدود الخطأ في نموذج الانحدار الخطى البسيط (١-١) غير طبيعي

عندما يكون  $H_0$  صحيح فإن  $r$  قيمة لمتغير عشوائي  $R_r$  له توزيع احتمالى. القيم الحرجة  $C_\alpha$  من  $R_r$  معطاه في الجدول في ملحق (٦) عند مستويات معنوية مختلفة. منطقة الرفض  $R_r \leq C_\alpha$  إذا وقعت  $r$  في منطقة الرفض نرفض  $H_0$ .

#### مثال (٢-٤)

بمطابق جدول (٢-٦) البواقي المعيارية المرتبة  $d_{(i)}$  الخاصة بالمثل (٢-٣) مع قيم  $Z_{(i)}, p_{(i)}$  حيث استخدم برنامج Mathematica في حساب قيم  $Z_{(i)}$ .

جدول (٢-٦)

$d_{(i)}$	$p_{(i)}$	$Z_{(i)}$
-1.4937	0.0510204	-1.63504
-1.42239	0.132653	-1.11394
-0.935205	0.214286	-0.791639
-0.51932	0.295918	-0.536176
-0.287661	0.377551	-0.311919
0.0985392	0.459184	-0.102491
0.300514	0.540816	0.102491
0.330198	0.622449	0.311919
0.472805	0.704082	0.536176
0.888689	0.785714	0.791639
1.27489	0.867347	1.11394
1.29264	0.94898	1.63504

والمطلوب اختبار فرض العدم:

$H_0$ : توزيع حدود الخطأ طبيعيه

ضد الفرض البديل:

$H_1$ : توزيع حدود الخطأ غير طبيعي

الحل

من البيانات في جدول (٦-٢) نحصل على صيغة  $r$  التالية :

$$r = \frac{\sum d_{(i)} z_{(i)} - \frac{\sum d_{(i)} \sum z_{(i)}}{n}}{\sqrt{\left[ \sum d_{(i)}^2 - \frac{(\sum d_{(i)})^2}{n} \right] \left[ \sum z_{(i)}^2 - \frac{(\sum z_{(i)})^2}{n} \right]}}$$

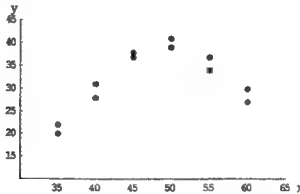
$$= \frac{9.74961}{\sqrt{(10)(9.87237)}} = 0.981243 .$$

لمستوى معنوية  $\alpha = 0.05$  فإن  $C_{0.05} \approx 0.918$  والمستخرجه من الجدول في ملحق (٦) عند  $n = 10$  وذلك لعدم وجود قيمة لـ  $C_{0.05}$  عند  $n = 12$ . منطقة الرفض  $R_r < 0.918$ . وبما أن  $r$  تقع في منطقة القبول نقبل  $H_0$ . أى أن حدود الخطأ تتبع التوزيع الطبيعي.

### (٢-٣) اختبار خطية الانحدار

#### Test for linearity of regression

الآن سوف نقدم اختبار إحصائي لنقص التوفيق لنموذج الانحدار الخطي البسيط .  
نفترض الطريقة تحقق كل من الاعتدال والاستقلال وثبات التباين فقط هناك شك في  
وجود علاقة خط مستقيم بين  $x, y$  . فعلى سبيل المثال للبيانات الموضحة في شكل  
(٢-٢٢) فإن هناك انطباع بالعين المجردة على أن الخط المستقيم ليس كافي لتوفيق  
البيانات وأننا نحتاج إلى اختبار يقدر لنا ما إذا كان هناك نظام على شكل منحني في  
توزيع النقاط على الرسم .



شكل (٢-٢٢)

يحتاج اختبار نقص التوفيق إلى وجود مشاهدات متكررة على الاستجابة  $y$  وذلك  
على الأقل لمستوى واحد من  $x$  . لقد أوضح (Draper and Smith, 1981) أن تلك  
المشاهدات المتكررة لابد أن تكون تكرارات حقيقية ، فعلى سبيل المثال بفرض أننا  
نحاول إيجاد العلاقة بين الذكاء ( $Y$ ) وطول الشخص ( $x$ ) . المشاهدات المتكررة  
يمكن الحصول عليها إذا أمكننا قياس ذكائين منفصلين لشخصين وبالضبط عدد نفس  
الطول ، في بعض الأحيان عندما يتم أخذ قراءتين للذكاء على نفس الشخص ، عند  
مستوى خاص من  $x$  ، فإن هذا لا يعتبر مشاهدات متكررة على الاستجابة  $Y$  التي  
نحتاج إليها في اختبار نقص التوفيق . بفرض أننا أخذنا عينة عشوائية من  $n$  من  
المشاهدات وذلك باستخدام  $m$  من القيم المختلفة من  $x$  ، ليكن  $x_1, x_2, \dots, x_m$   
بحيث أن العينة تحتوي  $n_1$  قيمة مشاهدة من المتغير العشوائي  $Y_1$  المقابل لـ  
 $x_1$  و  $n_2$  قيمة مشاهدة من المتغير العشوائي  $Y_2$  المقابل لـ  $x_2$  و  $n_m$  قيمة مشاهدة

من المتغير العشوائي  $Y_m$  المقابل لـ  $x_m$  من الضروري أن  $n = \sum_{i=1}^k n_i$ . سوف نعرف  $y_{ij}$  ليمثل القيمة رقم  $j$  من المتغير العشوائي  $Y_i$  عند  $x_i$  حيث  $i = 1, 2, \dots, k$  و  $j = 1, 2, \dots, n_i$  و  $y_i = \sum_{j=1}^{n_i} y_{ij}$  و  $\bar{y}_i = y_i / n_i$  وعلى ذلك عندما  $n_4 = 3$  فإن القياسات على  $Y_4$  تقابل  $x = x_4$  وسوف نعرف هذه المشاهدات بالرموز  $y_{41}, y_{42}, y_{43}$  وعلى ذلك  $y_4 = y_{41} + y_{42} + y_{43}$ . يعتمد الاختبار على تجزئة مجموع مربعات البواقي إلى جزئين كالتالي :

$$SSE = SSPE + SSLF,$$

حيث  $SSPE$  هو مجموع المربعات الذي يرجع إلى الخطأ الخالص (الصافي)  $pure\ error$  أي الاختلاف بين قيم  $y$  داخل قيمة معطاة من  $x$ . أما  $SSLF$  فهو مجموع مربعات نقص التوفيق أي أن  $SSPE$  يعكس الاختلاف العشوائي أو خطأ التجربة بينما  $SSLF$  يعتبر مقياس للاختلاف المنتظم الناتج من وجود حدود من درجات عليا ويجب أن نتذكر أنه في اختبار النموذج الخطي فلننا نفترض أن  $SSLF$  غير موجود وبالتالي فإن مجموع مربعات البواقي بالكامل يرجع إلى أخطاء عشوائية. للحصول على التجزئة السابقة لمجموع مربعات البواقي  $SSE$  فلننا نعرف أن الباقي رقم  $ij$  يعرف كالتالي :

$$y_{ij} - \hat{y}_i = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i). \quad (1-2)$$

بتربيع طرفي (1-2) والجمع على كل  $i$  نحصل على:

$$\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^m n_i (\bar{y}_i - \hat{y}_i)^2 \quad (2-2)$$

حيث الحد:

$$2 \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)(\bar{y}_i - \hat{y}_i) = 0$$

الحد الايسر من (2-2) هو مجموع مربعات البواقي  $SSE$ . الجزئين على الجانب الايمن يقيسان الخطأ الخالص ونقص التوفيق حيث مجموع مربعات الخطأ الخالص هو:

$$SSPE = \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^m \left[ \sum_{j=1}^{n_i} (y_{ij}^2) - \frac{(\sum y_{ij})^2}{n_i} \right]$$

والذي نحصل عليه بحساب مجموع المربعات المصحح للملاحظات المتكررة عند كل مستوى من  $x$  ثم الجمع على كل المستويات  $m$  من  $x$ . إذا كان فرض ثبات التباين متحقق فإن  $SSPE$  يعتبر مقياس للخطأ الخالص ولا يعتمد على النموذج وذلك لأن الاختلاف في قيم  $y$  عند كل مستوي من  $x$  هو الاختلاف الوحيد الذي يستخدم في حساب  $SSPE$ . ولأنه يوجد  $n_i - 1$  درجة حرية للخطأ الخالص عند كل مستوى  $x_i$  فإن العدد الكلي من درجات الحرية المرتبطة بمجموع مربعات الخطأ الخالص هو:

$$\sum_{i=1}^m (n_i - 1) = n - m.$$

مجموع مربعات نقص التوفيق هو :

$$SSLF = \sum_{i=1}^k n_i (\bar{y}_i - \hat{y}_i)^2 ,$$

والذي يمثل مجموع الانحرافات المربعة المرجحة بين  $\bar{y}_i$  ، عند كل مستوى من  $x$  ، والقيمة المقدرة المقابلة . عندما تقترب القيم المقدرة  $\hat{y}_i$  من  $\bar{y}_i$  المقابل لها فإن هذا يعتبر دليل قوي أن نموذج الانحدار خطي . وعندما تنحرف  $\hat{y}_i$  بدرجة كبيرة من  $\bar{y}_i$  فهذا يعني أن النموذج ليس خطي. يوجد  $m - 2$  من درجات الحرية يرتبط بـ  $SSLF$  وذلك لوجود  $m$  من مستويات  $x$  ودرجتين حرية تم فقدهم لأن هناك معلمتين لابد من تقديرهم وذلك للحصول على  $\hat{y}_i$  . في عملية الحساب عادة يتم الحصول على  $SSLF$  بطرح  $SSPE$  من  $SSE$  الإحصاء الذي يعتمد عليه الاختبار هو:

$$F = \frac{SSLF / (m - 2)}{SSPE / (n - m)} = \frac{MSLF}{MSPE} .$$

الحسابات المطلوبة لاختبار الفرض في مشكلة الانحدار بقياسات متكررة على الاستجابة يمكن تلخيصها كما هو موضح في جدول (٢-٧)

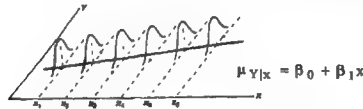
جدول (٧-٢)

S.O.V	df	SS	Ms	F
الاتحدار	1	SSR	SSR	$\frac{SSR}{MSE}$
الخطأ	n - 2	SSE	$MSE = SSE / (n - 2)$	
نقص التوفيق	m - 2	SSE - SSPE	$MSLF = \frac{SSE - SSPE}{m - 2}$	$\frac{MSLF}{MSPE}$
الخطأ الخالص	n - m	SSPE	$MSPE = \frac{SSPE}{n - m}$	
الكلية	n - 1			

القيمة المتوقعة لـ MSPE هي  $\sigma^2$  والقيمة المتوقعة لـ MSLF هي :

$$E(MSLF) = \sigma^2 + \frac{\sum_{i=1}^k n_i [\mu_{Y|X_i} - \beta_0 - \beta_1 x_i]^2}{m - 2} \quad (٣-٢)$$

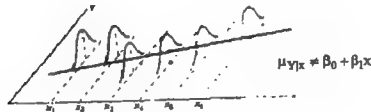
عندما يكون نموذج الاتحدار الخطي صحيح (فرض العدم  $H_0$ ) فإن  $\mu_{Y|X_i} = \beta_0 + \beta_1 x_i$  وبالتالي فإن الحد الثاني في (٣-٢) يصبح مساوياً للصفر والنتيجة أن  $E(MSLF) = \sigma^2$ . يوضح شكل (٢٣-٢) نقاط العينة للنموذج الصحيح حيث  $\mu_{Y|X_i}$  تقع على خط مستقيم ولا يوجد نقص في التوفيق عند فرض النموذج (١-١). وعلى ذلك اختلاف نقاط العينة حول خط الاتحدار هو خطأ خالص ناتج من الاختلاف الناتج من مشاهدات متكررة.



شكل (٢٣-٢)

عندما  $\mu_{Y|x_i} \neq \beta_0 + \beta_1 x_i$  فإن  $E(MSLF) > \sigma^2$  (الفرض البديل  $H_1$ ). يوضح شكل (٢-٢٤) نقاط العينة عندما يكون النموذج غير صحيح حيث يتضح أن  $\mu_{Y|x_i}$  لا تقع على الخط المستقيم وعلى ذلك نقص التوفيق الناتج من اختيار النموذج الخطي (١-١) يمثل جزء كبير من الاختلاف حول خط الانحدار بالإضافة إلى الخطأ الخالص . عندما يكون فرض العدم صحيح  $\mu_{Y|x_i} = \beta_0 + \beta_1 x_i$  فإن الإحصاء  $F$  يتبع توزيع  $F$  بدرجات حرية  $n-m$  و  $m-2$  . إذا كانت قيمة  $F$  المحسوبة أقل من قيمة  $F$  الجدولية فهذا يدل على أن هناك ثقة كبيرة في عدم نقص التوفيق وعندئذ نختبر الفرضية  $H_0: \beta_1 = 0$  باستخدام الإحصاء :

$$F = \frac{MSR}{MSE}$$



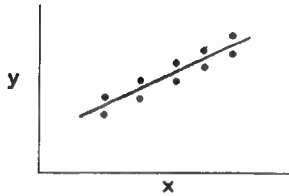
شكل (٢-٢٤)

وإذا كانت  $F$  المحسوبة أكبر من الجدولية فإننا نرفض فرض العدم  $H_0: \beta_1 = 0$  . لسوء الحظ وفي بعض الأحيان فإن رفض فرض العدم  $H_0: \beta_1 = 0$  يعطينا الضمان أن النموذج كافي لمعادلة الانحدار . اقترح (Box and Wetz 1973) أن القيمة المحسوبة من  $F$  لا بد أن تكون على الأقل أربعة أو خمس مرات القيمة الجدولية وذلك حتى يصبح نموذج الانحدار مفيد في التنبؤ . لشكل الانتشار المعطى في شكل



٢-٢٥) وعند قبول فرض العدم أن  $\mu_{Y|X_i} = \beta_0 + \beta_1 x_i$  ورفض فرض العدم  $H_0 : \beta_1 = 0$  فإن معادلة الانحدار المقترنة سوف تكون :

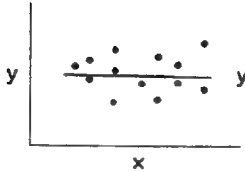
$$\hat{y} = b_0 + b_1 x$$



شكل (٢-٢٥)

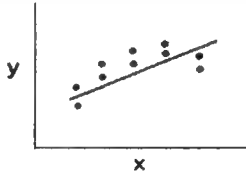
شكل الانتشار المعطى في شكل (٢-٢٦) وعند قبول فرض العدم أن النموذج هو  $\mu_{Y|X_i} = \beta_0 + \beta_1 x_i$  وقبول فرض العدم  $H_0 : \beta_1 = 0$  فإن معادلة الانحدار المقترنة سوف تكون :

$$\hat{y} = \bar{y}$$



شكل (٢٦-٢)

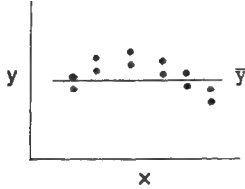
لشكل الانتشار المعطى في شكل (٢٧-٢) وعند رفض فرض العدم أن النموذج هو  $\mu_{y|x_i} = \beta_0 + \beta_1 x_i$  ورفض فرض العدم  $H_0: \beta_1 = 0$  فإننا نحاول مع النموذج  $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$  . أما إذا كان شكل الانتشار غير ذلك فلا بد من عمل تحويلات على قيم  $x$  أو قيم  $y$  أو كلاهما (وعلى الأكثر يستخدم التحويل لقيم  $y$ ) . هذا وهناك عدة طرق لتحويل البيانات سوف نتناولها في البند (٢-٥)



شكل (٢٧-٢)

لشكل الانتشار والمعطى في شكل (٢٨-٢) وعند رفض فرض العدم أن النموذج هو  $\mu_{y|x_i} = \beta_0 + \beta_1 x_i$  نقبل فرض العدم  $H_0: \beta_1 = 0$  نحاول مع النموذج

$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$  أما إذا كان الانتشار غير ذلك فابننا نلجأ إلى التحويلات .



شكل (٢-٢٨)

مثال (٢-٥)

توضح التجربة على نوع معين من البلاستيك أن هناك علاقة بين صلابة المواد المشكلة من البلاستيك  $Y$  مقاسه بوحدات برنيل ، والوقت المنعمر بعد انتهاء عملية التشكيل ( $x$ ) . صنعت ست عشرة عجينة من البلاستيك وشكلت وحدة اختبار واحدة من كل عجينة وخصصت كل وحدة اختبار إلى أحد أربعة مستويات زمنية محددة سلفاً ، وقيمت الصلابة بعد انقضاء الوقت المخصص ، والنتائج  $x$  الوقت المنعمر بالساعات ،  $y$  الصلابة مقاسه بوحدات برنيل معطاة في جدول (٢-٨) . افترض أن نموذج الانحدار (١-١) الخطي مناسب . والمطلوب استخدام اختبار  $F$  لتحديد ماذا كان هناك نقص توفيق لدالة انحدار خطية لم لا ؟

جدول (٢-٨)

$x_i$	16	16	16	16	24	24	24	24
$y_i$	199	205	196	200	218	220	215	223
$x_i$	32	32	32	32	40	40	40	40
$y_i$	237	234	235	230	250	248	253	246

الحل

الآن لاختبار نقص التوفيق أي اختبار فرض العدم :

$$H_0 : \mu_{Y|x} = \beta_0 + \beta_1 x$$

ضد الفرض البديل

$$H_1 : \mu_{Y|x} \neq \beta_0 + \beta_1 x .$$

نتبع الخطوات التالية:

مجموع مربعات الخطأ الخالص عند  $x=16$  هو :

$$(199)^2 + (205)^2 + (196)^2 + (200)^2 - \{(199 + 205 + 196 + 200)^2 / 4\} \\ = 42 .$$

بدرجات حرية  $(n_1 = 4 - 1 = 3)$

مجموع مربعات الخطأ الخالص عند  $x=24$  هو :

$$(218)^2 + (220)^2 + (215)^2 + (223)^2 - \{(218 + 220 + 215 + 223)^2 / 4\} \\ = 34 .$$

بدرجات حرية  $(n_2 = 4 - 1 = 3)$

بنفس الطريقة يمكن حساب مجموع مربعات الخطأ الخالص للقيم الباقية من  $x$  كما هو موضح في جدول (٩-٢).

جدول (٩-٢)

$x$	$\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$	درجات حرية
16	42	3
24	34	3
32	26	3
40	26.75	3

جدول تحليل التباين معطى في جدول (١٠-٢)

جدول (١٠-٢)

S.O.V	df	SS	MS	F
الانحدار	1	5297.51	5297.51	506.506
الخطأ	14	146.425	10.4589	--
مُصَوِّر التوفيق	2	17.675	8.8375	0.823689
الخطأ الخالص	12	128.75	10.72912	--

ومن جدول (١٠-٢) فإن قيمة F الخاصة بقصور التوفيق غير معنوية لأنها أقل من الواحد الصحيح .

مثال (٦-٢)

درست فعالية (جبر) تجريبي جديد في تخفيض استهلاك الجازولين في 12 محاولة استخدمت فيها عربة نقل خفيفة مجهزة بهذا الجبر حيث  $x$  في جدول (١١-٢) للسرعة الثابتة (بالميل في الساعة) لعربة الاختبار و  $y$  الأميال المقطوعة لكل جالون .

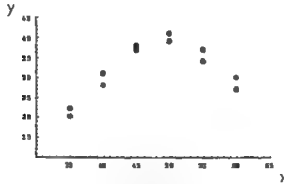
جدول (١١-٢)

x	y
35	22
35	20
40	28
40	31
45	37
45	38
50	41
50	39
55	34
55	37
60	27
60	30

فهل معادلة الخط المستقيم تلائم البيانات المعطاة في جدول (١١-٢) ؟

الحل

شكل الانتشار للبيانات في جدول (١١-٢) معطاة في شكل (٢٩-٢)



شكل (٢٩-٢)

نوجد أولاً معادلة الخط المستقيم المقدرة على افتراض أنها تلائم البيانات حيث :

$$\bar{y} = \frac{\sum y}{n} = \frac{384}{12} = 32, \quad \bar{x} = \frac{\sum x}{n} = \frac{570}{12} = 47.5$$

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$= \frac{18530 - \frac{(570)(384)}{12}}{27950 - \frac{(570)^2}{12}}$$

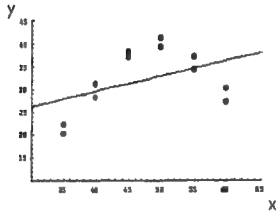
$$= \frac{290}{875} = 0.331429,$$

$$b_0 = \bar{y} - b_1\bar{x} = 32 - (0.331429)(47.5) \\ = 16.2571 .$$

معادلة الانحدار المقترحة سوف تكون على الشكل :

$$\hat{y} = 16.2571 + 0.331429x .$$

والممثلة بيانيا مع شكل الانتشار في شكل (٣٠-٢)



شكل (٣٠-٢)

جدول تحليل التباين معطى في جدول (١٢-٢).

جدول (١٢-٢)

S.O.V	df	SS	MS	F
الانحدار	1	96.1143	96.1143	2.32224
الخطأ	10	413.886	41.3886	--
الكلي	11	510	--	--

بما أن قيمة F المحسوبة (2.32224) أقل من القيمة الجدولية  $F_{0.05}[1,10] = 4.96$  فإننا نقبل فرض العدم  $H_0: \beta_1 = 0$  . والآن نختبر البواقي باستخدام رسم البواقي من معادلة الخط المستقيم المقدرة :

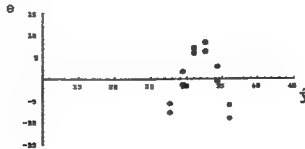
$$\hat{y} = 16.2571 + 0.331429x$$

نوجد  $e_i, d_i, r_i$  لكل قيم  $x_i$  كما هو معطى في جدول (١٣-٢)

جدول (١٣-٢)

$x_i$	$y_i$	$\hat{y}_i$	$e = y_i - \hat{y}_i$	$d_i$	$r_i$
35	22	27.8571	-5.8571	-0.9104	-0.9437
35	20	27.8571	-7.8571	-1.2213	-1.2658
40	28	29.5143	-1.5143	-0.2354	-0.2447
40	31	29.5143	1.4857	0.2309	0.2401
45	37	31.1714	5.8286	0.9059	0.9448
45	38	31.1714	6.8286	1.0614	1.1069
50	41	32.8286	8.1714	1.2701	1.3287
50	39	32.8286	6.1714	0.9592	1.0035
55	34	34.4857	-0.4857	-0.0755	-0.0792
55	37	34.4857	2.5143	0.3908	0.4101
60	27	36.1429	-9.1429	-1.4212	-1.4961
60	30	36.1429	-6.1429	-0.9548	-1.0052

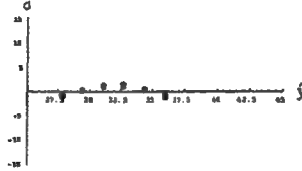
للبيانات في جدول (١٣-٢) والخاصة بالمثال (٦-٢) يوضح شكل (٣١-٢) رسم  $e_i$  مقابل  $\hat{y}_i$



شكل (٣١-٢)

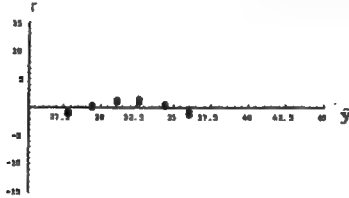


للبيانات في جدول (١٣-٢) والخاصة بالمثل (٦-٢) يوضح شكل (٣٢-٢) رسم  $d_i$  مقابل  $\hat{y}_i$



شكل (٣٢-٢)

وعند استخدام بواقي ستيوينجت نحصل على نفس الرسم ولكن مع اختلاف في مقياس الرسم كما يتضح من شكل (٣٣-٢)



شكل (٣٣-٢)

ومن ملاحظة الرسم البياني نرى بأنه يشبه  $\cap$  مما يدل على أن هناك معادلة من درجة ثانية سوف تكون أكثر ملائمة للبيانات . أي أن النموذج الخطي (١-١) لا

يلائم البيانات والذي يوضحه شكل الانتشار في شكل (٢-٣٠). ويمكن عمل اختبار لنقص التوفيق للتأكد كما يلي :

الآن لاختبار نقص التوفيق أي اختبار فرض العدم :

$$H_0 : \mu_{Y|x_i} = \beta_0 + \beta_1 x_i$$

ضد الفرض البديل:

$$H_1 : \mu_{Y|x_i} \neq \beta_0 + \beta_1 x_i$$

نتبع الخطوات التالية:

مجموع مربعات الخطأ الخالص عند  $x=35$  هو :

$$(22)^2 + (20)^2 - \{(22+20)^2 / 2\}$$

$$= 2$$

بدرجات حرية  $(n_1 = 2 - 1 = 1)$

مجموع مربعات الخطأ الخالص عند  $x=40$  هو :

$$(28)^2 + (31)^2 - \{(28+31)^2 / 2\}$$

$$= 4.5$$

بدرجات حرية  $(n_2 = 2 - 1 = 1)$

بنفس الطريقة يمكن حساب مجموع مربعات الخطأ الخالص للقيم الباقية من  $x$  كما هو موضح في جدول (٢-١٤) .

جدول (٢-١٤)

$x$	$n_i \sum_{j=1} (y_{ij} - \bar{y}_i)^2$	درجات الحرية
35	2	1
40	4.5	1
45	0.5	1
50	2	1
55	4.5	1
60	4.5	1

جدول تحليل التباين معطى في جدول (٢-١٥).

### جدول (٢-١٥)

S.O.V	df	SS	MS	F
الانحدار	1	96.1143	96.1143	2.32224
الخطأ	10	413.886	41.3886	--
قصور التوفيق	4	395.886	98.0714	32.9905
الخطأ الخالص	6	18	3	--

بما أن قيمة  $F$  المحسوبة لقصور التوفيق (32.9905) تزيد عن القيمة الجدولية  $F_{0.05}[4,6] = 4.53$  فإننا نرفض فرض العدم وبناء على ذلك فإن معادلة الخط المستقيم غير ملائمة للبيانات ويمكن استخدام معادلة من الدرجة الثانية.

### (٢-٤) تحويلات إلى الخط المستقيم

#### Transformations to a Straight Line

إن ضرورة اقتراح نموذج بسديل لنموذج الانحدار الخطي  $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$  يرجع إما إلى اعتبارات نظرية أو من الخبرة السابقة أو من اختبار رسوم البواقي أو من اختبار نقص جودة التوفيق. في كل حالة يكون من الضروري وضع نموذج معاملته يمكن تقديرها بسهولة. مجموعة خاصة من تلك النماذج يمكن تعريفها عن طريق ما يعرف بالدوال القابلة للتحويل إلى خطية  $\text{intrinsically linear}$  أو  $\text{transformably linear}$ .

**تعريف:** تسمى الدالة التي تربط  $x$  مع  $y$  بالدالة القابلة للتحويل إلى خطية إذا أمكن إجراء تحويلة على  $x$  (أو) تحويلة على  $y$  بحيث يمكن التعبير عن الدالة كالاتي  $y' = \beta_0 + \beta_1 x'$  حيث  $x'$  المتغير المستقل المحول و  $y'$  المتغير التابع المحول.

يعطي جدول (٢-١٦) بعض الدوال القابلة للتحويل إلى خطية. في كل حالة فإن التحويل المناسبة أما التحويل اللوغاريتمية (سواء للأساس 10 أو اللوغاريتم

الطبيعي للأساس  $e$  ) أو تحويلة المعكوس reciprocal . التمثيل البياني لتلك الدوال معطى في شكل (٢-٣٤) .

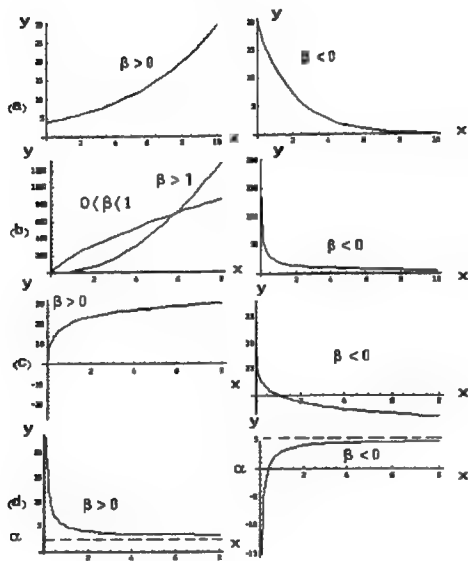
جدول (٢-١٦)

شكل علاقة الانحدار	التحويل الملائمة	الدالة التي تربط العلاقة بين $x, y$
$y' = \ln(\alpha) + \beta x$	$y' = \ln(y)$	a) الأسية $y = \alpha e^{\beta x}$
$y' = \log(\alpha) + \beta x'$	$y' = \log(y), x' = \log(x)$	b) القوى $y = \alpha x^{\beta}$
$y = \alpha + \beta x'$	$x' = \log(x)$	c) $y = \alpha + \beta \log(x)$
$y = \alpha + \beta x'$	$x' = \frac{1}{x}$	d) المعكوس $y = \alpha + \beta \cdot \frac{1}{x}$

عندما يوضح شكل انتشار  $y$  مقابل  $x$  أن العلاقة على شكل منحني فإننا نكون قادرين على مقارنة سلوك المشاهدات على الرسم مع واحد من المنحنيات المعطاة في شكل (٢-٣٤) واستخدام الشكل الخطي المحول للدالة وذلك لتوفيق البيانات .

يتضح من جدول (٢-١٦) أنه بالنسبة لعلاقة الدالة الأسية فإن  $y$  فقط التي تم تحويلها لتحقيق الخطية بينما في علاقة دالة القوى فإن كل  $x, y$  تم تحويلهما . ولأن المتغير  $x$  موجود في الأس في العلاقة الأسية فإن  $y$  تزيد (عندما  $\beta > 0$ ) أو تقل (عندما  $\beta < 0$ ) بسرعة أكبر وذلك بالمقارنة لنموذج القوى . ولكن خلال فترة قصيرة من قيم  $x$  يكون من الصعوبة التمييز بين الدالتين . هناك أمثلة للسوال التي لا يمكن التعبير عنها في صورة خطية مثل:

$$y = \alpha + \gamma e^{\beta x} \quad \text{أو} \quad y = \alpha + \gamma x^{\beta}$$



شكل (٢-٣٤)

الدالة المقابلة للتحويل إلى خطية تؤدي مباشرة إلى نماذج إحداد خطية ومعالمها يمكن تقديرها بسهولة باستخدام طريقة المربعات الصغرى العادية .

تعريف : نموذج الاتحاد الذي يربط  $Y$  بـ  $x$  يعتبر قابل للتحويل إلى خطية إذا أمكن إجراء تحويله على  $Y$  و  $(x \text{ أو } x)$  بحيث يمكن كتابته على الصورة :

$$Y' = \beta_0 + \beta_1 x' + \varepsilon'$$

يمكن تعريف نماذج الاتحاد القابلة للتحويل إلى خطية والمقابلة للدوال المعطاة في جدول (١٦-٢) كالآتي :

$$Y = \alpha e^{\beta x} \cdot \varepsilon \quad (a)$$

حيث حد الخطأ  $\varepsilon$  مضروب في  $\alpha e^{\beta x}$  . وعلى ذلك :

$$\ln(Y) = Y' = \beta_0 + \beta_1 x' + \varepsilon'$$

حيث :

$$x' = x, \quad \beta_0 = \ln(\alpha), \quad \beta_1 = \beta, \quad \varepsilon' = \ln(\varepsilon)$$

$$Y = \alpha x^{\beta} \cdot \varepsilon \quad (b)$$

حيث حد الخطأ  $\varepsilon$  مضروب في  $\alpha x^{\beta}$  وعلى ذلك :

$$\log(Y) = Y' = \beta_0 + \beta_1 x' + \varepsilon'$$

حيث :

$$x' = \log(x), \quad \beta_0 = \log(\alpha), \quad \beta_1 = \beta, \quad \varepsilon' = \log(\varepsilon).$$

$$Y = \alpha + \beta \log(x) + \varepsilon \quad (c)$$

حيث :

$$x' = \ln(x).$$

$$Y = \alpha + \beta \cdot \frac{1}{x} + \varepsilon \quad (d)$$

حيث :

$$x' = 1/x.$$

ومما يجدر الإشارة إليه أن نموذج الانحدار الأسّي على الصورة  $Y = \alpha e^{\beta x} + \varepsilon$  لا يعتبر نموذج انحدار قابل للتحويل إلى خطي وبنفس الشكل نموذج القوى الذي على الصورة  $Y = \alpha x^{\beta} + \varepsilon$ . يتطلب النموذج (a) والنموذج (b) تحويلة على  $Y$  وهذا بدوره يؤدي إلى تحويلة على حد الخطأ  $\varepsilon$ . في الحقيقة إذا كان  $\varepsilon$  يتبع التوزيع الطبيعي اللوغاريتمي بمتوسط  $e^{\sigma^2/2}$  وتباين يساوي  $\tau^2$  مستقل عن  $x$  فإن النماذج المحورة في (a) و (b) سوف تحقق كل الفروض الخاصة بنموذج الانحدار الخطي (١-١)، وهذا يؤدي إلى أن كل الاستدلالات على المعالم للنموذج المحول و التي تعتمد على هذه الفروض سوف تكون صحيحة. عندما يكون تباين حد الخطأ  $\sigma^2$  صغير فإن  $\mu_{Y|x} \approx \alpha e^{\beta x}$  في (a) أو  $\mu_{Y|x} \approx \alpha x^{\beta}$  في (b).

الميزة الأساسية لنموذج الانحدار القابل للتحويل إلى خطي هو أن المعلمتين  $\beta_0, \beta_1$  في النموذج المحول يمكن تقديرهما باستخدام طريقة المربعات الصغرى العادية وذلك بالتعويض عن  $x', y'$  في صيغة كل من  $b_0, b_1$  كالتالي :

$$b_1 = \frac{\sum x'_i y'_i - \frac{\sum x'_i \sum y'_i}{n}}{\sum (x'_i)^2 - \frac{(\sum x'_i)^2}{n}} \quad (٢-٤)$$

$$b'_0 = \frac{\sum y'_i}{n} - b_1 \frac{\sum x'_i}{n}$$

المعالم في نموذج الانحدار الغير خطي الأصلي يمكن تقديرها لأنها تكون دالة في  $b_0, b_1$  وذلك عند الضرورة.

عند تحليل البيانات المحولة يجب الأخذ في الاعتبار النقاط التالية :

- تقدير  $\beta_0, \beta_1$  كما في (٢-٤) ثم إعادة التحويل للحصول على تقديرات للمعالم الأصلية لا يكفي استخدام طريقة المربعات الصغرى والتي تطبق مباشرة على النموذج الأصلي فعلى سبيل المثال باستخدام للنموذج الأسّي يمكن تقدير  $\alpha, \beta$  بطريقة المربعات الصغرى وذلك بتفسير  $\Sigma (y_i - \hat{\alpha} e^{\hat{\beta} x_i})^2$  التقديرات الناتجة سوف تختلف:  $\hat{\alpha} \neq e^{b_0}, \hat{\beta} \neq b_1$ .

• إذا كان النموذج المختار غير قابل للتحويل إلى نموذج خطي (النماذج الغير خطية) فإن الطريقة في (٢-٤) لا يمكن استخدامها . وبدلاً من ذلك يمكن استخدام طريقة المربعات الصغرى ( أو طرق أخرى للتوفيق ) والتي تطبق على النموذج الغير محول. وعلى ذلك للنموذج  $Y = \alpha e^{\beta x} + \varepsilon$  فإن طريقه المربعات الصغرى سوف تؤدي إلى تصغير  $\sum (y_i - \hat{\alpha} e^{\hat{\beta} x})^2$  وبأخذ التفاضلات الجزئية بالنسبة لكل من  $\hat{\alpha}$  ,  $\hat{\beta}$  , نحصل على معادلتين طبيعيتين غير خطيتين في  $\hat{\alpha}$  ,  $\hat{\beta}$  . والتي يمكن حلها باستخدام أي طريقة من طرق التكرارات iterative procedure . وسوف نتناول النماذج الغير خطية بتفصيل أكثر في الفصل العاشر.

#### مثال (٢-٧)

لأزواج القياسات في جدول (٢-١٧) أوجد معادلة الانحدار المقدرة تحت فرض النموذج الأسى حيث  $\mu_{y|x} = \gamma \delta^x$

#### جدول (٢-١٧)

x	1	2	3	4	5	6	7
y	304	341	393	457	548	670	882

#### الحل

يوضع  $y'_i = \ln y_i$  فإن  $\sum y'_i = 43.243148$

$$n = 7, \quad \sum x_i = 28, \quad \sum x_i^2 = 140$$

$$\sum x_i y'_i = 177.85134, \quad \bar{x} = 4, \quad \frac{\sum y'_i}{n} = 6.1775926,$$

$$b_1 = \frac{177.85134 - \frac{(28)(43.243148)}{7}}{140 - \frac{(28)^2}{7}} = \frac{4.87875}{28}$$



١٥٦.

$$= 0.174241 .$$

$$b_0 = 6.1775926 - (0.174241)(4) \\ = 5.4806286 .$$

معادلة الانحدار المقدرة هي :

$$\hat{y} = 5.4806286 + 0.174241x.$$

وعلى ذلك :

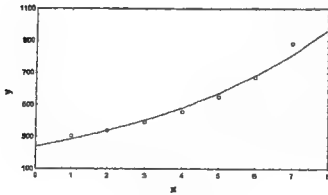
$$\ln d = b_1 = 0.174241 , \ln c = b_0 = 5.4806286,$$

$$d = \exp(b_1) = 1.1903424, c = \exp(b_0) = 239.99752.$$

وبالتالي فإن منحنى الانحدار المقدر بالمربعات الصغرى هو :

$$\hat{y} = c d^x \\ = (239.99752)(1.1903424)^x.$$

والتمثيل البياني لها موضح في شكل (٣٥-٢)



شكل (٣٥-٢)

مثال (٢-٨)

لأزواج القياسات في جدول (٢-١٨) لوجد معادلة الانحدار المقدرة تحت فرض نموذج القوى.

جدول (٢-١٨)

x	600	600	600	600	500	500	500	500	400	400	400	400
y	2.35	2.65	3.0	3.6	6.4	7.8	9.8	16.5	21.5	24.5	26.0	33.0

الحل

$$\begin{aligned}
 n &= 12, \quad \sum \ln x_i = 74.412, \quad \sum \ln y_i = 26.22601, \\
 \sum \ln x_i^2 &= 461.75874, \quad \sum (\ln x_i)(\ln y_i) = 160.84601, \\
 \sum \ln y_i^2 &= 67.74609. \\
 b_1 &= \frac{160.84601 - \frac{(74.412)(26.22601)}{12}}{461.75874 - \frac{(74.412)^2}{12}} = \frac{-1.78146}{0.329915} \\
 &= -5.3996, \\
 b_0 &= \frac{26.22601 - (-5.3996)(74.412)}{12} \\
 &= 35.6684.
 \end{aligned}$$

معادلة الانحدار المقدرة هي :

$$\hat{y} = 35.6684 - 5.3996x.$$

وعلى ذلك :

$$\ln \hat{\alpha} = b_0 = 35.6684$$

أي أن :

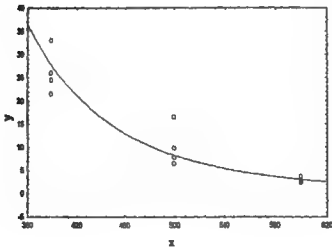
$$\hat{\alpha} = \exp(b_0) = 3.094491530.10^{15},$$

$$\hat{\beta} = b_1 = -5.3996.$$

والمعادلة الأساسية المقدرة هي :

$$\hat{y} = \hat{\alpha} x^{\hat{\beta}} = 3.094491530.10^{15} \cdot x^{-5.3996}.$$

والتمثيل البياني لها موضح في شكل (٣٦-٢) مع شكل الانتشار .



شكل (٣٦-٢)

## مثال (٢-٩)

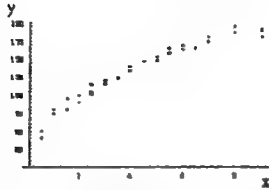
يعطي جدول (٢-١٩) عدد الساعات التي يقضيها 30 طالب في الدراسة خارج المدرج في الأسبوع ( $x$ ) والدرجات التي حصلوا عليها في مادة الإحصاء ( $y$ ) حيث الدرجة النهائية 200. هل يمكن تمثيل البيانات بمعادلة خط مستقيم ؟

جدول (٢-١٩)

$x$	$y$	$x^2$	$xy$
0.5	40	0.25	20
0.5	50	0.25	25
1	75	1	75
1	80	1	80
1.5	90	2.25	120
1.5	95	2.25	142.5
2	100	4	200
2	90	4	180
2.5	114	6.25	285
2.5	103	6.25	257.5
2.5	101	6.25	252.5
3	116	9	348
3	120	9	360
3.5	123	12.25	430.5
4	138	16	552
4	133	16	532
4.5	146	20.25	657
5	152	25	760
5	147	25	735
5.5	157	30.25	863.5
5.5	164	30.25	902
6	167	36	1002
6	162	36	972
6.5	164	42.25	1066
7	173	49	1211
7	179	49	1253
8	186	64	1488
8	193	64	1544
9	190	81	1710
9	180	81	1620
127	Σy	729	19643.5

الحل

شكل الانتشار موضح في شكل (٢-٣٧)



شكل (٢-٣٧)

بفرض أن نموذج الاتجاه الخطي البسيط (١-١) يمثل البيانات في جدول (٢-١٩) فإن :

$$n = 30, \quad \bar{x} = \frac{\sum x}{n} = \frac{127}{30} = 4.23333, \quad \bar{y} = \frac{\sum y}{n} = \frac{3918}{30} = 130.6$$

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{19643.5 - \frac{(127)(3918)}{30}}{729 - \frac{(127)^2}{30}} \\ = \frac{3057.3}{191.367} = 15.9761,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 130.6 - (15.9761)(4.23333) = 62.9677 .$$

معادلة خط الاتجاه المقدرة سوف تكون على الشكل :

$$y = 62.9677 + 15.9761x .$$

لما جدول تحليل التباين فهو كما في جدول (٢-٢٠) .

## جدول (٢-٢٠)

S.O.V.	df	SS	MS	F
الاتحداد	1	48843.8	48843.8	384.45
الخطأ	28	3557.36	127.048	—
الكل	29	52401.2	—	—

F الجدولية هي  $F_{0.05}[1,28] = 4.2$  . وبما إن F المحسوبة اكبر نرفض فرض العدم  $H_0 : \beta_1 = 0$

ولمعرفة مدى توفر شروط فروض التحليل نتبع ما يلي :

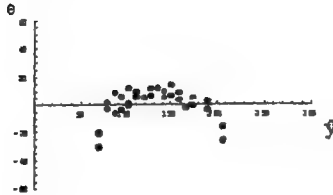
١- نحسب قيم البواقي  $e_i$  و البواقي المصايرة  $d_i$  وبواقي سيودنت  $t_i$  والنتائج معطاة في

جدول (٢-٢١)

جدول (٢-٢١) .

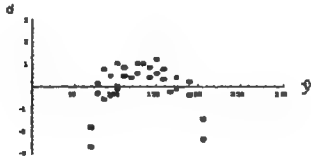
$x_i$	$y_i$	$\hat{y}_i$	$e_i$	$d_i$	$t_i$
0.5	40	70.9557	-30.9557	-2.7463	-2.9048
0.5	50	70.9557	-20.9557	-1.8591	-1.9664
1	75	78.9438	-3.9438	-0.3498	-0.3663
1	80	78.9438	1.0561	0.0937	0.0981
1.5	80	86.9318	-6.9318	-0.6149	-0.6385
1.5	95	86.9318	8.0681	0.7157	0.7431
2	100	94.9199	5.0800	0.4506	0.4647
2	90	94.9199	-4.9199	-0.4364	-0.4500
2.5	114	102.9080	11.0919	0.9840	1.0091
2.5	103	102.9080	0.0919	0.0081	0.0083
2.5	101	102.9080	-1.9080	-0.1692	-0.1735
3	116	110.8960	5.1039	0.4528	0.4624
3	120	110.8960	9.1039	0.8076	0.8248
3.5	123	118.8841	4.1158	0.3651	0.3719
4	138	126.8722	11.1277	0.9972	1.0042
4	133	126.8722	6.1277	0.5436	0.5530
4.5	146	134.8603	11.1396	0.9882	1.0053
5	152	142.8483	9.1516	0.8119	0.8271
5	147	142.8483	4.1516	0.3683	0.3752
5.5	157	150.8364	6.1635	0.5468	0.5585
5.5	164	150.8364	13.1635	1.1678	1.1930
6	167	158.8245	8.1754	0.7253	0.7440
6	162	158.8245	3.1754	0.2817	0.2889
6.5	164	166.8125	-2.8125	-0.2495	-0.2573
7	173	174.8006	-1.8006	-0.1597	-0.1659
7	179	174.8006	4.1993	0.3725	0.3870
8	186	190.7767	-4.7767	-0.4237	-0.4485
8	193	190.7767	2.2232	0.1972	0.2087
9	190	206.7529	-16.7529	-1.4862	-1.6140
9	180	206.7529	-26.7529	-2.3734	-2.5775

٢- نرسم البواقي  $e_i$  مقابل  $\hat{y}_i$  والنتائج معطاة في شكل (٢-٣٨)

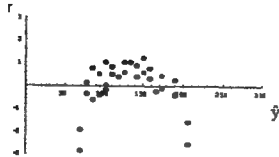


شكل (٢-٣٨)

من ملاحظة رسم البواقي في شكل (٢-٣٨) نرى بأنه على شكل منحنى  $\cap$  مما يدل على أن النموذج الخطي لا يلائم البيانات. نفس الشيء في شكل (٢-٣٩) عند رسم البواقي المعيارية  $d_i$  مقابل  $\hat{y}_i$ . أما شكل (٢-٤٠) فنحصل عليه عند رسم بواقي ستودنت مقابل رسم  $\hat{y}_i$ .



شكل (٢-٣٩)



شكل (٢-٤٠)

٣- بما أن هناك تكرار لقيم  $x$  فإنه يمكن عمل اختبار لنقص التوفيق كما يلي : يتم حساب مجموع مربعات الخطأ الخالص من جدول (٢-٢٢) .

جدول (٢-٢٢)

$x$	مجموع مربعات الخطأ الخالص	درجات الحرية
0.5	50	1
1	12.5	1
1.5	112.5	1
2	50	1
2.5	98	2
3	8	1
3.5	0	0
4	12.5	1
4.5	0	0
5	12.5	1
5.5	24.5	1
6	12.5	1
6.5	0	0
7	18	1
8	24.5	1
9	50	1
	485.5	14



جدول تحليل التباين معطى في جدول (٢-٢٣) .

جدول (٢-٢٣)

S.O.V.	df	SS	MS	F
الانحدار	1	48843.8	48843.8	384.45
الخطأ	28	3557.36	127.048	--
قصور التوفيق	14	3071.86	219.418	6.327
الخطأ الخالص	14	485.5	34.6786	

وبما أن قيمه F المحسوبة لنقص المطابقة (6.327) أكبر من القيمة F الجدولية  $F_{0.05}[14,14] \approx 2.53$  عند مستوى معنوية  $\alpha = 0.05$  لذا فإن النموذج الخطي لا يلائم البيانات بل أن هناك معادلة أخرى قد تلائم البيانات والذي يتضح من خلال شكل الانتشار (٢-٣٧) وعلى ذلك يمكن المحاولة مع تحويله على  $x$  حيث  $x' = \sqrt{x}$  .

النموذج الخطي سيصبح :

$$Y = \beta_0' + \beta_1'x' + \varepsilon'$$

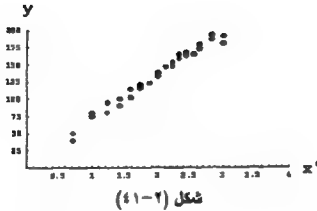
ومن جدول (٢-٢٤) يتم حساب كل من :

$$x', y, x'^2, x'y$$

جدول (٢-٢٤)

x	y	x'	x' <sup>2</sup>	x'y
0.5	40	0.7071	0.5000	28.2842
0.5	50	0.7071	0.5000	35.3553
1	75	1	1	75
1	80	1	1	80
1.5	80	1.2247	1.4999	97.9795
1.5	95	1.2247	1.4999	116.3507
2	100	1.4142	2.0000	141.4213
2	90	1.4142	2.0000	127.2792
2.5	114	1.5811	2.5000	180.2498
2.5	103	1.5811	2.5000	162.8572
2.5	101	1.5811	2.5000	159.6950
3	116	1.7320	2.9999	200.9178
3	120	1.7320	2.9999	207.8460
3.5	123	1.8708	3.5	230.1119
4	138	2	4	276
4	133	2	4	266
4.5	146	2.1213	4.4999	309.7127
5	152	2.2360	5.0000	339.8823
5	147	2.2360	5.0000	328.7019
5.5	157	2.3452	5.5	368.1976
5.5	164	2.3452	5.5	384.6140
6	167	2.4494	5.9999	409.0647
6	162	2.4494	5.9999	396.8173
6.5	164	2.5495	6.4999	418.1196
7	173	2.6457	7.0000	457.7149
7	179	2.6457	7.0000	473.5894
8	186	2.8284	8.0000	526.0874
8	193	2.8284	8.0000	545.8864
9	190	3	9	570
9	180	3	9	540

لازواج القيم (x',y) المعطى في جدول (٢-٢٤) فإن شكل الانتشار موضح في شكل (٢-٤١).



الآن يتم حساب القيم التالية وللإيجاد معادلة الانحدار المقدرة:

$$SXY = 820.011 ,$$

$$SXX = 13.1153 ,$$

$$b'_1 = \frac{SXY}{SXX} = \frac{820.011}{13.1153} = 62.5235 ,$$

$$b'_0 = \bar{y} - b'_1 \bar{x} = 130.6 - (62.5235)(1.94837) = 8.78096 .$$

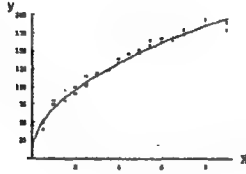
معادلة الخط المستقيم المقدرة هي :

$$\hat{y} = 8.78096 + 62.5235x'$$

والممثلة بيانيا في شكل (٢-٢) مع شكل الانتشار للبيانات الأصلية.

والتي تصبح على الشكل :

$$\hat{y} = 8.78096 + 62.5235\sqrt{x}$$



شكل (٢-٤٢)

جدول تحليل التباين للبيانات المحولة معطى في جدول (٢-٢٥).

جدول (٢-٢٥)

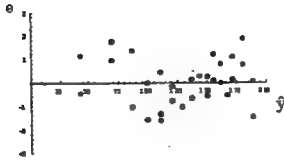
S.O.V	df	SS	MS	F
الاتحاد	1	51270	51270	1269
الخطأ	28	1131.25	40.4017	--
الكل	29	52401.2	--	--

بما أن قيمة  $F$  المحسوبة تزيد عن قيمة  $F$  الجدوليه  $F_{05}(1,28) = 4.2$  فإننا نرفض فرض العدم  $H_0: \beta_1 = 0$ . البواقي  $e_i$  والبواقي المعيارية  $d_i$  وبواقي ستودنت  $t_i$  معطاة في جدول (٢-٢٦).

جدول (٢-٢٦)

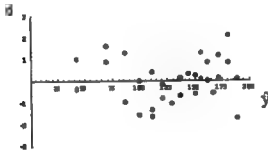
$x'_i$	$y_i$	$\hat{y}_i$	$e_i$	$d_i$	$r_i$
0.707107	40.	52.9917	-12.9917	-2.04393	-2.21801
0.707107	50	52.9917	-2.99173	-0.470676	-0.510763
1.	75	71.3044	3.69558	0.581409	0.613511
1.	80	71.3044	8.69558	1.36804	1.44357
1.22474	80	85.3563	-5.35625	-0.842677	-0.87535
1.22474	95	85.3563	9.64375	1.51721	1.57604
1.41421	100	97.2025	2.79751	0.44012	0.452768
1.41421	90	97.2025	-7.20249	-1.13314	-1.1657
1.58114	114	107.639	6.36076	1.00071	1.02328
1.58114	103	107.639	-4.63924	-0.729872	-0.746329
1.58114	101	107.639	-6.63924	-1.04452	-1.06808
1.73205	116	117.075	-1.07478	-0.16909	-0.172299
1.73205	120	117.075	2.92522	0.460213	0.468947
1.87083	123	125.752	-2.75165	-0.432906	-0.440411
2.	138	133.828	4.17211	0.656381	0.667672
2.	133	133.828	-0.827888	-0.130248	-0.132489
2.12132	146	141.413	4.58674	0.721613	0.734817
2.23607	152	148.588	3.41233	0.536847	0.547815
2.23607	147	148.588	-1.58767	-0.249782	-0.254886
2.34521	157	155.411	1.58852	0.249915	0.255781
2.34521	164	155.411	8.58852	1.3512	1.38291
2.44949	167	161.932	5.06846	0.797399	0.819184
2.44949	162	161.932	0.0684572	0.0107701	0.0110643
2.54951	164	168.185	-4.18514	-0.658431	-0.67944
2.64575	173	174.202	-1.2025	-0.189184	-0.196218
2.64575	179	174.202	4.7975	0.754771	0.782836
2.82843	186	185.624	0.37598	0.0591513	0.0620889
2.82843	193	185.624	7.37598	1.16043	1.21806
3.	190	196.351	-6.35135	-0.999231	-1.06377
3.	180	196.351	-16.3514	-2.57249	-2.73864

رسم البواقي  $e_i$  مقابل  $\hat{y}_i$  معطاة في شكل (٢-٤٣)



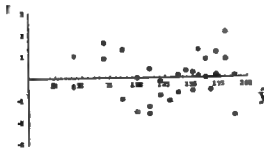
شكل (٤٣-٢)

أيضاً رسم البواقي المعيارية  $d_i$  مقابل  $\hat{y}_i$  معطى في شكل (٤٤-٢) .



شكل (٤٤-٢)

وأخيراً رسم بواقي ستودنت  $t_i$  مقابل  $\hat{y}_i$  معطى في شكل (٤٥-٢)



شكل (٤٥-٢)

يتضح من شكل (٢-٤٣) شكل (٢-٤٤) وشكل (٢-٤٥) أن  $e_i$  تتوزع توزيعاً عشوائياً حول الصفر مما يدل على أن النموذج للمحول أكثر ملاءمة من النموذج الأول . ومما يؤكد ذلك أيضاً أن  $R^2$  للنموذج الثاني  $\left( \frac{51270}{52401.2} = 0.979 \right)$  أكبر من  $R^2$  للنموذج الأول  $\left( \frac{48843.8}{52401.16} = 0.932 \right)$  وأن MSE للنموذج الثاني أصغر من قيمته في النموذج الأول وعليه فإن التحويلة  $x' = \sqrt{x}$  مناسبة .

#### (٢-٥) اكتشاف وتصحيح عدم ثبات التباين

##### (٢-٥-١) مقدمه

يطلق على تحقق الفرض  $\text{Var}(e_i) = \text{Var}(Y_i) = \sigma^2$  ثبات التباين لحندود الاخطاء ، أو اختصاراً ثبات التباين - تجانس التباين ، homoscedasticity بينما مخالفة هذا الفرض يسمى عدم ثبات التباين heteroscedasticity . يعتبر ثبات التباين المطلوب الأساسي لتحليل الانحدار وفي حالة عدم تحققه فإن مقدرات المربعات الصغرى العادية لمعالم نموذج الانحدار الخطي البسيط (١-١) تمتلك صفة عدم التحيز ولكن لن تمتلك صفة أقل تباين ، أي لن تكون أفضل تقدير خطي غير متحيز. أيضاً عدم ثبات التباين قد يؤثر على معامل التحديد  $R^2$  واختبارات الفروض وفترات الثقة.

غالباً السبب العام لعدم ثبات التباين هو أن المتغير التابع  $Y_i$  يتبع توزيع احتمالي حيث التباين دالة في المتوسط . فعلى سبيل المثال إذا كان المتغير التابع متغير قابل للعد ، على سبيل المثال يتبع تقريباً توزيع بواسون ، فإن التباين  $\sigma_i^2$  للمتغير  $Y_i$  يساوي  $E(Y_i)$  .

في هذه الحالة فإن تحويلات لتثبيت التباين سوف تكون مفيدة. وعلى ذلك إذا كان المتغير  $Y_i$  يتبع توزيع بواسون فإنه يمكن استخدام التحويل  $Y' = \sqrt{Y}$  وذلك لأن التباين للجذر التربيعي لمتغير عشوائي يتبع بواسون يكون مستقل عن المتوسط. عندما  $Y_i = \frac{m_i}{n_i}$  يمثل نسبة من أعداد  $n_i$  و  $m_i$  فإن تباين  $Y_i$  من المحتمل يقترب من  $E(Y_i) \cdot (1 - E(Y_i)) / n_i$ . هذه الحالة يمكن اكتشافها من رسم البواقي مقابل  $\hat{Y}$  حيث نحصل على القوس المزوج كما في شكل (٢-٨). في هذه الحالة فإن استخدام التحويلة  $Y' = \sin^{-1}(\sqrt{Y})$  سوف تكون مفيدة في تثبيت التباين. عندما  $Y_i$  هو المتوسط  $\sum_{i=1}^{n_i} z_i / n_i$  وذلك لمتغيرات عشوائية  $Z_1, Z_2, \dots, Z_{n_i}$  لها تباين ثابت ، فإن  $\sigma_i^2 \propto n_i^{-1}$ .

في بعض الأحيان فإن تباين  $Y_i$  تختلف باختلاف  $x_i$  أو  $\frac{1}{x_i}$  أو  $\frac{1}{x_i^2}$ .... الخ. أيضا من الممكن أن يختلف تباين  $Y$  مع متغيرات غير موجودة في النموذج. فعلى سبيل المثال عند استخدام أجهزة معملية مختلفة أو آلات مختلفة قد تؤثر على التباين. في مثل هذه الحالات ، يكون من المفيد رسم البواقي مقابل المتغير المستقل وكل متغير يتوقع أن يؤثر على التباين. هناك الكثير من الرسوم المقترحة مثل القيم المطلقة للبواقي أو الجذر التربيعي للبواقي أو لوغاريتم القيم المطلقة للبواقي أو البواقي المعيارية وذلك مقابل  $\hat{Y}$ . لمزيد من المعلومات عن طرق أخرى يمكن الرجوع إلى (1985) Carroll and Ruppert و Cook (1980) and Weisberg.

#### (٢-٥-٢) طرق تحليلية لاكتشاف عدم ثبات التباين

يوجد طرق عديدة لاختبار عدم تجانس التباين. سوف نقدم في الجزء التالي طريقتين .



### ١- طريقة جولد فيلد - كواندت

#### Coldfield-Quandt (1965)

يمكن استخدام هذه الطريقة في حالة وجود متغير مستقل (أو أكثر) حيث :  
ترتب المشاهدات وفقاً لأحد المتغيرات المستقلة ترتيب تصاعدي أو تنازلي ثم  
يُحذف 20% من المشاهدات من مركز السلسلة وليكن (c) وذلك يجعل الاختبار  
أكثر حساسية. يستخدم الجزء الأول من المشاهدات  $\frac{(n-c)}{2}$  في إيجاد معادلة  
الاتحدار المطلوبة والحصول على مجموع مربعات الخطأ  $SSE_1$  من جدول  
تحليل التباين. نكرر ما سبق في الخطوة التالية ولكن باستخدام المشاهدات الأخيرة  
وعندها أيضاً  $\frac{(n-c)}{2}$  وإجراء انحدار والحصول على مجموع مربعات  
الخطأ  $SSE_2$ . يستخدم اختبار جولد فيلد - كواندت للكشف عن نوعين من عدم  
ثبات التباين وهما :

(أ) عندما يكون تباين حد الخطأ دالة تناقصية في المتغير المستقل x حيث  
فرض العدم سوف يكون  $H_0$  : تباين حد الخطأ متجانس ضد الفرض  
البديل  $H_1$  : تباين حد الخطأ دالة تناقصية في المتغير x. وفي هذا  
الاختبار يستخدم الأحصاء F الذي يأخذ الصيغة التالية:

$$F = \frac{SSE_1 (n-c)/2}{SSE_2 (n-c)/2} = \frac{MSE_1}{MSE_2} \quad (٥-٢)$$

ومقارنة قيمة F المحسوبة بنظيرتها الجدولية بدرجات حرية الخطأ  
للعמוד والصف وإذا كانت قيمة F أكبر من نظيرتها الجدولية نرفض  
فرض العدم.

(ب) عندما يكون تباين حد الخطأ دالة تزايدية للمتغير المستقل x فإن فرض  
العدم يكون  $H_0$  : تباين حد الخطأ متجانس ضد الفرض البديل  $H_1$  :  
تباين حد الخطأ دالة تزايدية في x وفي هذا الاختبار يستخدم الأحصاء  
F على الصورة التالية :

$$F = \frac{SSE_2/(n-c)/2}{SSE_1/(n-c)/2}$$

وبمقارنة قيمة  $F$  المحسوبة بنظيرتها الجدولية بدرجات حرية الخطأ للعمود والصف وإذا كانت قيمة  $F$  أكبر من نظيراتها الجدولية نرفض فرض العدم .

### مثال (١٠-٢)

البيانات المعطاة في جدول (٢٧-٢) تمثل درجات اختبار القبول ودرجات اختبار التفاضل والتكامل لعشرة من طلبة الجامعة ، مع أمل أن يكون هؤلاء العشرة عينة عشوائية من مجتمع الطلبة في الجامعة والمطلوب اختبار تجانس التباين.

جدول ( ٢٧-٢ )

الطلاب	الدرجة في امتحان القبول $x$	الدرجة في امتحان التفاضل $y$
1	39	65
2	43	74
3	21	52
4	64	82
5	57	92
6	47	74
7	28	73
8	75	98
9	34	56
10	52	75

### الحل

وحيث أن عدد أزواج المشاهدات  $n=10$  فلنأخذ الأربعة أزواج الأولى من المشاهدات المرتبة وفقاً للمتغير  $x$  ونستخدمها في إيجاد جدول تحليل التباين كما هو موضح في جدول (٢٨-٢) وذلك باستخدام برنامج Mathematica

ونلك للحصول على  $SSE_1$  ونفس الطريقة نحصل على  $SSE_2$  لزوج المشاهدات الاربعة الاخيرة المرتبة وفقاً للمتغير  $x$  كما هو موضح في جدول (٢٩-٢) ثم نحسب قيمة  $F$ .

جدول (٢٨-٢)

Source	df	SS	MS	F
Regression	1	28.6409	28.6409	0.242351
Residual	2	236.359	118.18	-
Total	3	265	-	-

نتائج جدول (٢٨-٢) هي :

$$R^2 = 0.108079 \quad \text{و} \quad SSE_1 = 236.359$$

$$\hat{y} = 49.3674 + 0.39779x.$$

جدول (٢٩-٢)

Source	df	SS	MS	F
Regression	1	174.443	174.443	2.486
Residual	2	140.307	70.1535	-
Total	3	314.75	-	-

نتائج جدول (٢٩-٢) هي :

$$R^2 = 0.554227 \quad \text{و} \quad SSE_2 = 140.307$$

$$\hat{y} = 39.3138 + 0.765101x.$$

بالتعويض في المعادلة (٥-٢) وعلية فقيمة  $F$  للمثال (١٠-٢) تكون:

$$F = \frac{236.359/2}{140.307/2} = 1.6845.$$

وبما أن قيمة  $F$  المحسوبه أقل من القيمة الجدولية  $(F_{0.05}(2,2) = 19)$  فإننا نقبل فرضى العدم وهو ثبات التباين .

## ٢- اختبار معامل سبيرمان

يعتمد هذا الاختبار على التقييم المطلقة للأخطاء. يحسب معامل سبيرمان لارتباط الرتب وذلك وفق الصيغة التالية:

$$r_{s,x} = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)},$$

حيث  $d$  الفرق بين رتب القيم المطلقة للباقي ورتب المتغير المستقل. وكلما كانت قيمة معامل الارتباط عالية وقريبة من الواحد الصحيح دل ذلك على وجود علاقة قوية بين الأخطاء والمتغير المستقل وبالتالي وجود مشكلة عدم ثبات التباين.

أيضاً يمكن اختبار مدى معنوية ثبات التباين في العينة تحت البحث وذلك بحساب الانحراف المعياري لمعامل سبيرمان وذلك وفق الصيغة التالية:

$$S(r_{e.x}) = \frac{1}{\sqrt{n-1}}.$$

ثم نحسب قيمة من قيم الأحصاء  $Z \sim (0,1)$  من الصيغة التالية:

$$z = \frac{r_{e.x}}{S(r_{e.x})} = \frac{r_{e.x}}{1/\sqrt{n-1}} = r_{e.x} \sqrt{n-1}.$$

لمستوى معنوي  $\alpha$  وإذا كانت القيمة المحسوبة تقع في الفترة  $z_{\alpha/2} < Z < z_{\alpha/2}$  حيث  $z_{\alpha/2}$  تستخرج من جدول التوزيع التقياسي الطبيعي في ملحق (٥)، نقبل فرض العدم أن هناك ثبات في التباين وغير ذلك نقبل الفرض البديل والذي يعني وجود مشكلة عدم ثبات التباين مع ملاحظة أن اختبار  $Z$  يفضل استخدامه عندما تكون  $n > 20$ .

مثال (٢-١١)

للبيانات المعطاه في جدول (٢-٣٠) و جدول (٢-٣١) أختبر فيما إذا كانت العلاقة الخطية بين  $x, Y$  خاضعة لفرضية ثبات التباين أو لعمه.

جدول (۲-۳۰)

$y_i$	$\hat{y}_i$	$e_i$
75.3	75.39182813	-0.09182813
85	81.83972321	3.16027679
87.97	84.4537342	3.5162658
82	83.8437987	-1.8437987
85.9	85.32507189	0.57492811
81.4	80.44558373	0.95441627
81.5	76.52456604	4.97543396
84.9	74.69475839	10.20524161
75.9	78.09297337	-2.19297337
57.5	73.82342122	-16.32342122
70	97.87232717	-27.87232717
127.5	123.9253086	3.5746914
139.5	131.8544769	7.6455231
148	133.5100175	14.4899825
173.6	158.77877955	14.8212205
174.6	199.0781397	-24.4781397
185.8	176.815475	8.984525

## جدول (٢-٣١)

$x_i$	الرتبة	$ e_i $	الرتبة	$d_i$	$d_i^2$
96	3	0.09182813	1	2	4
103.4	7	3.16027679	6	1	1
106.4	9	3.5162658	7	2	4
105.7	8	1.8437987	4	4	16
107.4	10	0.57462811	2	8	64
101.8	6	0.95441627	3	3	9
97.3	4	4.97543396	9	-5	25
95.2	2	10.20524161	12	-10	100
99.1	5	2.19297337	5	0	0
94.2	1	16.32342122	15	-14	196
121.8	11	27.87232717	17	-6	36
151.7	12	3.5740914	8	4	16
160.0	13	7.6455231	10	3	9
162.7	14	14.4899825	13	1	1
191.7	15	14.8212205	14	1	1
237.95	17	24.4781397	16	1	1
212.4	16	8.984525	11	5	25

الحل

$$\sum_{i=1}^n d_i^2 = 508,$$

$$r_{e.x} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} = 1 - \frac{6(508)}{17(288)} = 1 - 0.622549019 = 0.37745098 .$$

الفرضية المطلوب اختبارها :

$$H_0 : \rho_{e.x} = 0 ,$$

$$H_1 : \rho_{e.x} \neq 0 ,$$

(حيث  $\rho_{e.x}$  هو معامل الارتباط للمجتمع).

$$S(r_{e.x}) = \frac{1}{\sqrt{n-1}} = \frac{1}{\sqrt{17-1}} = 0.25 ,$$

$$z = \frac{r_{e.x}}{S(r_{e.x})} = \frac{0.37745098}{0.25} = 1.50980392 ,$$

لمستوى معنوية  $\alpha=0.05$

وبما أن  $Z$  المحسوبة تقع في منطقة القبول :

$$-z_{0.025} \leq Z_0 \leq z_{0.025} .$$

حيث  $z_{0.025}=1.96$  نستخرج من جدول التوزيع الطبيعي القياسي في ملحق (٥)

ومنه نقبل فرضية عدم الفاتلة بأن بيانات العينة لا تعاني من وجود مشكلة عدم تجانس تباين الخطأ وبالتالي يمكن إيجاد معادله الانحدار المقدرة بأسلوب طريقة المربعات الصغرى العادية (OLS) .

وتجدر الإشارة هنا إلى أن الاختبار أعلاه قد يطبق في حالة وجود أكثر من متغير مستقل واحد في نموذج الانحدار الخطي .

### (٢-٥-٢) تصحيح عدم ثبات التباين

عندما لا يتحقق ثبات التباين لحدود الخطأ  $\varepsilon_i$  في نموذج الانحدار الخطي البسيط (١-١) فلا بد من اتخاذ إجراء وذلك لجعل تباينات  $\varepsilon_i$  (أو  $Y_i$  تقريباً) متساوية. يتم تصحيح عدم ثبات التباين بطريقتين :

١- الطريقة الأولى : عمل تحويل لقيم  $y$  .

٢- الطريقة الثانية : استخدام طريقة المربعات الصغرى المرجحة بدلاً من استخدام طريقة المربعات الصغرى العادية وذلك باستخدام وزن معين  $w_i$

لجعل التباين للأخطاء متجانس. سوف نتناول الطريقتين بشئ من التفصيل في الجزء التالي :

### ١- التحويلات

السؤال الآن كيف نحصل على تحويلة  $y$  المناسبة ؟ الجواب سوف نحصل عليه بإتباع الآتى :

لأي دالة  $f(Y)$  و  $Y$  والتي المشتقة الأولى  $f'(Y)$  لها موجودة ومتصلة والمشتقة الثانية  $f''(Y)$  منتهية فإننا نعلم من مبادئ التفاضل أن:

$$f(Y_i) - f(\eta_i) = (y_i - \eta_i)f'(\eta_i) + \frac{1}{2}(y_i - \eta_i)^2 f''(\theta)$$

حيث  $\theta$  تقع بين  $Y_i$  و  $\eta_i$  و  $\eta_i = E(Y_i)$  . عندما  $(y_i - \eta_i)^2$  صغيرة فإن:

$$f(Y_i) - f(\eta_i) \approx f'(\eta_i)(Y_i - \eta_i). \quad (٦-٢)$$

بترتيب طرفي (٦-٢) وأخذ التوقع فإننا نحصل على تقريب لتباين  $f(Y_i)$  على الشكل التالي:

$$\text{Var}(f(Y_i)) \approx (f'(\eta_i))^2 \sigma_i^2(\eta_i),$$

حيث  $\sigma_i^2(\eta_i)$  هو تباين المتغير العشوائي  $Y_i$  بمتوسط  $\eta_i$  . وعلى ذلك ، لإيجاد تحويلة مناسبة  $f$  لـ  $Y_i$  فلا بد من جعل  $\text{Var}(f(Y_i))$  تقريبا ثابت وذلك بحل المعادلة:

$$f'(\eta_i) = c/\sigma_i(\eta_i), \quad (٧-٢)$$

حيث  $c$  ثابت. تسمى التحويلة  $f$  بتحويلة تثبيت التباين. فلي سبل المثال :

بفرض الحالة عندما  $Y_i$  متغير قابل للعد ، فإن  $\sigma_i^2(\eta_i) \propto \eta_i$  ونحتاج  $f$  بحيث أن:

$$f'(\eta_i) = c / \eta_i^{\frac{1}{2}} \quad (٨-٢)$$

من الواضح ، عندما نختار  $c = \frac{1}{2}$  فإن  $f(\eta_i) = \eta_i^{\frac{1}{2}}$  تحل المعادلة (٨-٢)

وفي هذه الحالة فإن  $Y_i^{\frac{1}{2}}$  هي تحويلة تثبيت التباين . الآن بفرض أن  $Y_i = m_i/n_i$  والتي تمثل نسبة وعلى ذلك  $Y_i$  يتبع توزيع ذي الحدين. التباين لـ  $Y_i$  هو  $\text{Var}(Y_i) = n_i^{-1}\eta_i(1-\eta_i)$  حيث  $\eta_i = \mu_{Y|x_i}$  . وعلى ذلك فإننا نحتاج الى حل :



$$f'(\eta_i) = cn_i \frac{1}{2} \left/ \left( \eta_i \frac{1}{2} (1 - \eta_i) \right)^{\frac{1}{2}} \right. \quad (٩-٢)$$

وبإجراء التكامل لطرفي (٩-٢) بالنسبة لـ  $\eta_i$  نحصل على:

$$\begin{aligned} f(\eta_i) &= \int f'(\eta_i) d\eta_i = cn_i \frac{1}{2} \int \frac{d\eta_i}{\eta_i^{\frac{1}{2}} (1 - \eta_i)^{\frac{1}{2}}} \\ &= 2cn_i \frac{1}{2} \sin^{-1}(\sqrt{\eta_i}) \end{aligned}$$

والذي يعطي  $n_i \frac{1}{2} \sin^{-1}(\sqrt{Y_i})$  كتحويله تقريبية. كمثال آخر عندما  $\sigma_i = n_i$

فإن (٧-٢) تؤدي إلى  $f(\eta_i) = \log(n_i)$

يعطي جدول (٣٢-٢) ملخص لمعظم التحويلات التي تستخدم لتثبيت التباين .

جدول (٣٢-٢)

التحويلة	علاقة $\sigma^2$ مع $E(Y)$
$Y' = Y$ (لا يوجد تحويلة)	$\sigma^2 \propto \text{constant}$
(بيانات بواسون وتحويلة الجذر التربيعي) $Y' = \sqrt{Y}$	$\sigma^2 \propto E(Y)$
(بيانات ذي الحدين وتحويلة الزلوي) $Y' = \sin^{-1}(\sqrt{Y})$	$\sigma^2 \propto E(Y)(1 - E(Y))$
(اللوغاريتم) $Y' = \ln(Y)$	$\sigma^2 \propto [E(Y)]^2$
(تحويلة معكوس الجذر التربيعي) $Y' = Y^{-\frac{1}{2}}$	$\sigma^2 \propto [E(Y)]^3$
(المعكوس) $Y' = Y^{-1}$	$\sigma^2 \propto [E(Y)]^4$

في بعض الاحيان يمكن إستخدام خبره قبله أو اعتبارات نظريه وذلك للمساعدة في اختيار التحويله المناسبه ، أو كبدل الاستعانة برسم اليواقي.

### ب- المربعات الصغرى المرجحة

#### Weighted least squares:

بفرض أن  $\text{Var}(\varepsilon_i) = \sigma_i^2 = \sigma^2/w_i$  حيث  $w_i$  أوزان معروفة. يمكن تثبيت التباين بضرب طرفي نموذج الانحدار الخطي البسيط (١-١) بثابت  $c_i$  حيث  $c_i = \sqrt{w_i}$  كالآتي :

$$c_i Y_i = c_i \beta_0 + c_i \beta_1 x_i + c_i \varepsilon_i, \quad i = 1, 2, 3, \dots, n. \quad (10-2)$$

من الواضح أن النموذج (١٠-٢) يحقق ثبات التجانس. للحصول على المعالم  $\beta_0, \beta_1$  يمكن استخدام طريقة المربعات الصغرى المرجحة Weight Least Squares (WLS) والتي تؤدي إلى تصغير مجموع المربعات المرجحة .

$$\sum_{i=1}^n w_i (y_i - b_0 - b_1 x_i)^2. \quad (11-2)$$

التقديرات للمعالم  $\beta_0, \beta_1$  والتي نحصل عليها بتصغير (١١-٢) تسمى تقديرات المربعات الصغرى المرجحة. عندما  $w_i = 1$  فإن طريقة المربعات الصغرى المرجحة تصبح طريقة المربعات الصغرى الاعتيادية Ordinary Least Squares (OLS). تقديرات المربعات الصغرى المرجحة متوافرة في كثير من الحزم الاحصائية. المعادلات الطبيعية لطريقة المربعات الصغرى هي:

$$b_0 \sum w_i + b_1 \sum w_i x_i = \sum w_i y_i$$

$$b_0 \sum w_i x_i + b_1 \sum w_i x_i^2 = \sum w_i x_i y_i.$$

بحل المعادلتين السابقتين آنيا نحصل على تقديرات لمعالم  $\beta_0, \beta_1$  حيث

$$b_1 = \frac{\sum x_i y_i w_i - \frac{\sum x_i w_i \sum y_i w_i}{\sum w_i}}{\sum x_i^2 w_i - \frac{(\sum x_i w_i)^2}{\sum w_i}} \quad (12-2)$$

$$b_0 = \frac{\sum y_i w_i}{\sum w_i} - b_1 \frac{\sum x_i w_i}{\sum w_i} \quad (13-2)$$

في كثير من المشاكل فإن الأوزان يمكن تقديرها بسهولة. على سبيل المثال إذا كانت  $y_i$  مشاهدة في الحققة تمثل متوسط مشاهدات مأخوذة من عينة حجمها

$n_i$  عند  $x_i$  وإذا كانت كل المشاهدات الأصلية لها تباين ثابت  $\sigma^2$  فإن تباين  $Y_i$  هو  $\text{Var}(Y_i) = \text{Var}(\varepsilon_i) = \sigma^2/n_i$  ولذلك يكون الوزن هو  $n_i$ . في بعض الأحيان تباين  $Y_i$  يكون دالة في المتغير المستقل  $x$  ، فعلى سبيل المثال

$$w_i = \frac{1}{x_i} \quad \text{Var}(Y_i) = \text{Var}(\varepsilon_i) = \sigma^2 x_i$$

كوزن. أيضاً قد يكون  $\text{Var}(Y_i) = \sigma^2 x_i^2$  وعلى ذلك  $w_i = \frac{1}{x_i^2}$  والتي

تظهر كثيراً في الدراسات الأبحاث التي تعتمد على بيانات إحصائية تأخذ شكل البيانات المقطعية cross-section data حيث تشتمل مشاهدات البيانات المقطعية الخاصة بالمتغير التابع قد تختلف اختلافاً كبيراً من مستوى إلى آخر من مستويات المتغير المستقل . فعلى سبيل المثال في دراسة العلاقة بين دخل وأنفاق الأسر على مختلف السلع والخدمات نجد أن الأسر ذات الدخل المرتفعة تتمتع بمرونة كبيرة في الأنفاق ، في حين أن أنفاق الأسر ذات الدخل المنخفضة يقع عادة ضمن حدود ضيقة وعليه فإن التباين عند الدخل الكبير ، يكون أكبر من التباين عند قيم الدخل الصغيرة و هكذا نجد أن فرضية ثبات التباين تصبح عديمة الجدوى في مثل هذه الحالات وبالتالي يولج الباحث مشكلة عدم ثبات التباين والتي سوف نصحها في المثال التالي .. في كثير من المشاكل فإن الأوزان لا تكون معروفة في البداية ونحتاج إلى تقديرها بالاعتماد على نتائج مربعات الصغرى العادية وسوف نتناول هذه الحالة لاحقاً.

#### مثال (٢-١٢)

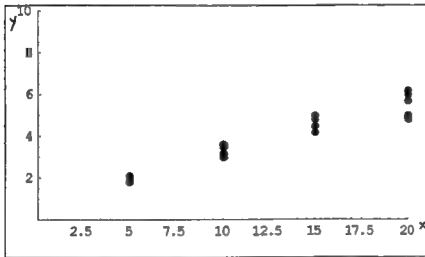
يعطي جدول (٢-٣٣) بيانات عن الدخل والإنفاق الشهري لعينة مكونة من 20 مشاهدة مقسمة إلى أربعة مجاميع وكل مجموعة بها خمس مشاهدات. العينة في كل مجموعة تم الحصول عليها بتقدير الإنفاق الشهري لخمس أسر عند نفس الدخل. أوجد تقديرات معالم نموذج الانحدار الخطي تحت فرض أن

$$\text{Var}(\varepsilon_i) = \sigma^2 x_i^2$$

جدول (٢-٣٣)

الدخل \$1000	الإنفاق الشهري \$ 1000					الجموعة
5.0	2.1	2.0	2.0	2.0	1.8	1
10.0	3.6	3.5	3.5	3.2	3.0	2
15.0	5.0	4.8	4.5	4.2	4.2	3
20.0	6.2	6.0	5.7	5.0	4.8	4

يوضح شكل (٢-٤٦) أن العلاقة بين الإنفاق والدخل الشهري علاقة خطية.



شكل (٢-٤٦)

البيانات اللازمة لحساب معادلة الانحدار المقترنة معطاة في جدول (٢-٣٤).

جدول (۲-۴)

x	y	$x^2$	xy
5	1.8	25	9
5	2	25	10
5	2	25	10
5	2	25	10
5	2.1	25	10.5
10	3	100	30
10	3.2	100	32
10	3.5	100	35
10	3.5	100	35
10	3.6	100	36
15	4.2	225	63
15	4.2	225	63
15	4.5	225	67.5
15	4.8	225	72
15	5	225	75
20	4.8	400	96
20	5	400	100
20	5.7	400	114
20	6	400	120
20	6.2	400	124
250	77.1	3750	1112

حیث:

$$\bar{y} = \frac{\sum y}{n} = \frac{77.1}{20} = 3.855, \quad \bar{x} = \frac{\sum x}{n} = \frac{250}{20} = 12.5$$

$$SXY = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$= 1112 - \frac{(250)(77.1)}{20}$$

$$= 148.25,$$

$$SXX = \sum x^2 - \frac{\sum x_i^2}{n}$$

$$= 3750 - \frac{(250)^2}{20} = 625,$$

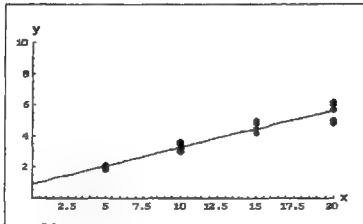
$$b_1 = \frac{SXY}{SXX} = \frac{148.25}{625} = 0.2372,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 3.855 - 0.2372(12.5) = 0.89 .$$

وعلى ذلك معادلة الانحدار المقدرة هي:

$$\hat{y} = 0.89 + 0.2372 x.$$

والعمثلة بيانياً في شكل (٤٧-٢) مع شكل الانتشار.



شكل (٤٧-٢)

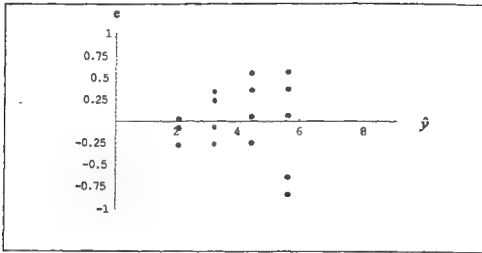
يعطي جدول (٢-٣٥) البواقي  $e_i$  والبواقي المعيارية  $d_i$  وبواقي ستودننت  $r_i$ .

والموضحة بيانيا في شكل (٢-٤٨) وشكل (٢-٤٩) وشكل (٢-٥٠) على التوالي.

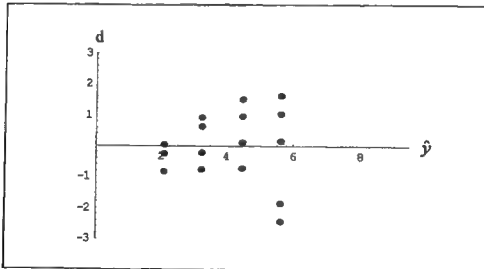
جدول (٢-٣٥)

x	y	$\hat{y}_i$	$e_i$	$d_i$	$r_i$
5	1.8	2.076	-0.276	-0.739905	-0.79786
5	2	2.076	-0.076	-0.203742	-0.219701
5	2	2.076	-0.076	-0.203742	-0.219701
5	2	2.076	-0.076	-0.203742	-0.219701
5	2.1	2.076	0.024	0.0643396	0.0693792
10	3	3.262	-0.262	-0.702374	-0.724443
10	3.2	3.262	-0.062	-0.166211	-0.171433
10	3.5	3.262	0.238	0.638034	0.658082
10	3.5	3.262	0.238	0.638034	0.658082
10	3.6	3.262	0.338	0.906116	0.934587
15	4.2	4.448	-0.248	-0.664842	-0.685733
15	4.2	4.448	-0.248	-0.664842	-0.685733
15	4.5	4.448	0.052	0.139402	0.143783
15	4.8	4.448	0.352	0.943647	0.973298
15	5	4.448	0.552	1.47981	1.52631
20	4.8	5.634	-0.834	-2.2358	-2.41093
20	5	5.634	-0.634	-1.69964	-1.83277
20	5.7	5.634	0.066	0.176934	0.190793
20	6	5.634	0.366	0.981179	1.05803
20	6.2	5.634	0.566	1.51734	1.63619

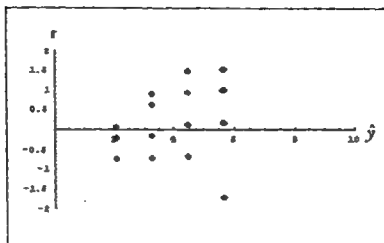




شکل (۴۸-۲)



شکل (۴۹-۲)



شكل (٢-٥٠)

يتضح من رسم البواقي في شكل (٢-٤٨) وشكل (٢-٤٩) وشكل (٢-٥٠) أن تباين البواقي غير ثابت وذلك لظهور الشكل القمعي المقروح من الأمام.

للتحقق أكثر من تحقق فرض عدم ثبات التباين نقوم بإجراء اختبار جولد فيلد - كواندت. ولإجراء هذا الاختبار نقوم بتقسيم المشاهدات إلى قسمين. القسم الأول يشمل الدخول من \$5.000 إلى \$10.000 والقسم الثاني يشمل الدخول العاليه من \$15.000 إلى \$20.000 ولايستبعد هنا مشاهدات من الوسط ومن بعد ذلك نقوم بتقدير معادلة الانحدار للقيم الصغيرة من  $x_i$  وأخرى للقيم الكبيرة من  $x_i$  حيث معادلة الانحدار المقدرة للقيم الصغيرة من الدخل هي:

$$\hat{y} = .600 + 0.276x.$$

$$R^2 = 0.94, \quad SSE_1 = 0.3.$$

ومعادلة الانحدار المقدرة للقيم الكبيرة من الدخل هي:

$$\hat{y} = 1.540 + 0.20x$$

$$R^2 = 0.55, \quad SSE_2 = 2.024.$$

وبمقارنة قيمة  $F$  المحسوبة ( $SSE_2 / SSE_1 = 6.7$ ) بالقيمة الجدولية  $F_{0.01}[8,8] = 6.03$  نجد أن قيمة  $F$  المحسوبة تزيد عن القيمة الجدولية وهذا يعني رفض فرض عدم وقبول الفرض البديل بعدم تجانس التباين. الآن باستخدام طريقة المربعات الصغرى المرجحة فإن البيانات اللازمة لحساب تقديرات المعامل  $\beta_1, \beta_0$  معطاة في جدول (٢-٣٦).

جنول (۲-۳۶)

x	y	$w = \frac{1}{x^2}$	xw	xyw	yw
5	1.8	0.04	0.2	0.36	0.072
5	2	$\frac{1}{25}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{25}$
5	2	$\frac{1}{25}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{25}$
5	2	$\frac{1}{25}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{25}$
5	2.1	$\frac{1}{25}$	$\frac{1}{5}$	0.42	0.084
10	3	$\frac{1}{100}$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{3}{100}$
10	3.2	0.01	0.1	0.32	0.032
10	3.5	$\frac{1}{100}$	$\frac{1}{10}$	0.35	0.035
10	3.5	$\frac{1}{100}$	$\frac{1}{10}$	0.35	0.035
10	3.6	0.01	0.1	0.36	0.036
15	4.2	$\frac{1}{225}$	$\frac{1}{15}$	0.28	0.0186667
15	4.2	$\frac{1}{225}$	$\frac{1}{15}$	0.28	0.018667
15	4.5	$\frac{1}{225}$	$\frac{1}{15}$	0.3	0.02
15	4.8	$\frac{1}{225}$	$\frac{1}{15}$	0.32	0.0213333
15	5	$\frac{1}{225}$	$\frac{1}{15}$	$\frac{1}{3}$	$\frac{1}{45}$
20	4.8	$\frac{1}{400}$	$\frac{1}{20}$	0.24	0.012
20	5	$\frac{1}{400}$	$\frac{1}{20}$	$\frac{1}{4}$	$\frac{1}{80}$
20	5.7	$\frac{1}{400}$	$\frac{1}{20}$	0.285	0.01425
20	6	$\frac{1}{400}$	$\frac{1}{20}$	$\frac{3}{10}$	$\frac{3}{200}$
20	6.2	$\frac{1}{400}$	$\frac{1}{20}$	0.31	0.0155

وعلى ذلك :

$$b_1 = \frac{\sum w_i x_i y_i - \frac{\sum w_i x_i \sum w_i y_i}{\sum w_i}}{\sum w_i x_i^2 - \frac{(\sum w_i x_i)^2}{\sum w_i}}$$

$$= 0.249487,$$

$$b_0 = \frac{\sum w_i y_i - b_1 \sum w_i x_i}{\sum w_i}$$

$$= 0.752923 .$$

وعلى ذلك معادلة الانحدار المقدرة سوف تكون:

$$\hat{y} = 0.752923 + 0.249487x .$$

يتم حساب جدول تحليل التباين من البيانات في جدول (٢-٣٧).

جدول (٢-٣٧)

w	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$	$w(y - \hat{y})^2$
0.04	1.8	2.00036	-0.200359	0.0401437	0.00160575
$\frac{1}{25}$	2	2.00036	-0.000358974	$1.28863 \times 10^{-7}$	$5.1545 \times 10^{-9}$
$\frac{1}{25}$	2	2.00036	-0.000358974	$1.28863 \times 10^{-7}$	$5.1545 \times 10^{-9}$
$\frac{1}{25}$	2	2.00036	-0.000358974	$1.28863 \times 10^{-7}$	$5.1545 \times 10^{-9}$
$\frac{1}{25}$	2.1	2.00036	0.099641	0.00992833	0.000397133
$\frac{1}{100}$	3	3.24779	-0.247795	0.0614023	0.000614023
0.01	3.2	3.24779	-0.0477949	0.00228435	0.0000228435
$\frac{1}{100}$	3.5	3.24779	0.252205	0.0636074	0.000636074
$\frac{1}{100}$	3.5	3.24779	0.252205	0.0636074	0.000636074
0.01	3.6	3.24779	0.352205	0.124048	0.00124048
$\frac{1}{225}$	4.2	4.49523	-0.295231	0.0871612	0.000387383
$\frac{1}{225}$	4.2	4.49523	-0.295231	0.0871612	0.000387383
$\frac{1}{125}$	4.5	4.49523	0.00476923	0.0000227456	$1.01091 \times 10^{-7}$
$\frac{1}{125}$	4.8	4.49523	0.304769	0.0928843	0.000412819
$\frac{1}{225}$	5	4.49523	0.504769	0.254792	0.00113241
$\frac{1}{400}$	4.8	5.74267	-0.942667	0.88862	0.00222155
$\frac{1}{400}$	5	5.74267	-0.742667	0.551554	0.00137888
$\frac{1}{400}$	5.7	5.74267	-0.0426667	0.00182044	$4.55111 \times 10^{-6}$
$\frac{1}{400}$	6	5.74267	0.257333	0.0662204	0.000165551
$\frac{1}{100}$	6.2	5.74267	0.457333	0.209154	0.000522884

جدول تحليل التباين معطى في جدول (٢-٣٨) حيث:

مجموع المربعات البواقي سوف تكون:

$$SSE = \sum w(y_i - \hat{y}_i)^2 = 0.0117659 .$$

مجموع المربعات الكلي سوف يكون:

$$\begin{aligned}
 SSYY &= \sum y_i^2 w_i - \frac{(\sum w_i y_i)^2}{\sum w_i} \\
 &= 0.307804 .
 \end{aligned}$$

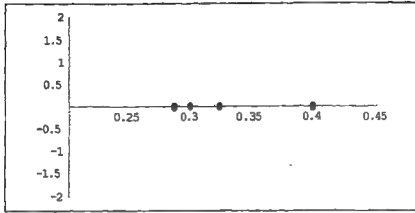
جدول تحليل التباين معطى في جدول (٣٨-٢).

جدول (٣٨-٢)

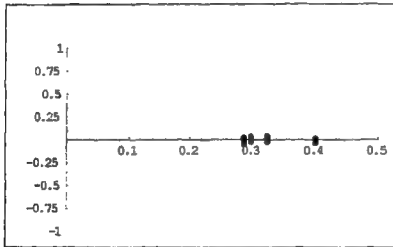
source	df	SS	MS	F
regression	1	0.29603783	0.29603783	452.89
residual	18	0.0117659	0.000653662	--
Total	19	0.307804	-	--

من جدول (٣٨-٢) وبما أن قيمة  $F$  المحسوبة تزيد عن القيمة الجدولية  $F_{0.05}[1,18]=4.41$  فإننا نقبل الفرض البديل أن  $H_1: \beta_1 \neq 0$ .

يعطى شكل (٥١-٢) رسم البواقي  $\sqrt{w}(y_i - \hat{y}_i)$  مقابل  $\sqrt{w_i}\hat{y}_i$  ، كما يعطى شكل (٥٢-٢) رسم البواقي  $\sqrt{w_i}(y_i - \hat{y}_i)$  مقابل  $\sqrt{w_i}x_i$  . يتضح من شكل (٥١-٢) وشكل (٥٢-٢) أن البواقي تنتشر حول الصفر وهذا يعني تجانس التباين.



شكل (٢-٥١)



شكل (٢-٥٢)

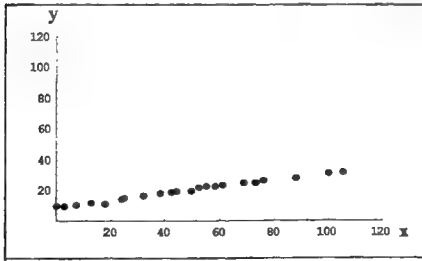
مثال (٢-١٣)

في دراسة للعلاقة بين نمو الفرع معينة من الاشجار  $Y$  مع الزمن  $x$  تم الحصول على البيانات المعطاة في جدول (٢-٣٩) .

جدول (٢-٣٩)

x	$n_i$	$\bar{y}_i$	$s_i$
0	5	10.20	0.83
3	5	10.40	0.54
7	5	10.60	0.54
13	6	12.50	0.83
18	5	12.00	1.41
24	4	15.00	0.82
25	6	15.17	0.76
32	5	17.00	0.72
38	7	18.71	0.74
42	9	19.22	0.84
44	10	20.00	1.26
49	19	20.32	1.00
52	14	22.07	1.20
55	11	22.64	1.76
58	9	22.78	0.84
61	14	23.93	1.16
69	10	25.50	0.98
73	12	25.08	1.94
76	9	26.67	1.23
88	7	28.00	1.01
100	10	31.67	1.42
106	7	32.14	2.28

شكل الانتشار موضح في شكل (٢-٥٣)



شكل (٢-٥٣)



والمطلوب إيجاد معادلة الانحدار المقدرة باستخدام طريقة المربعات الصغرى  
المرجحة حيث يعتبر الوزن للملاحظة رقم  $i$  هو  $w_i = n_i$ .

الحل

معادلة الانحدار المقدرة سوف تكون :

$$b_1 = \frac{\sum x_i y_i w_i - \frac{\sum x_i w_i \sum y_i w_i}{\sum w_i}}{\sum x_i^2 w_i - \frac{(\sum x_i w_i)^2}{\sum w_i}}$$

$$= 0.21733,$$

$$b_0 = \frac{\sum y_i w_i}{\sum w_i} - b_1 \frac{\sum x_i w_i}{\sum w_i}$$

$$= 9.97375 .$$

وعلى ذلك معادلة الانحدار المقدرة سوف سيكون :

$$\hat{y} = 9.97375 + 0.21733x .$$

لاختبار فرض العدم  $H_0 : \beta_1 = 0$  ضد الفرض البديل  $H_1 : \beta_1 \neq 0$

يتم حساب جدول تحليل التباين (٢-٤١) من جدول (٢-٤٠) حيث  
مجموع المربعات للبواقي المرجحة سوف تكون :

$$SSE = \sum w_i (y_i - \hat{y}_i)^2 = 74.3921 .$$

مجموع المربعات الكلي سيكون :

$$SYY = \sum y_i^2 w_i - \frac{(\sum y_i w_i)^2}{\sum w_i} = 6238.67 .$$

جدول (٢-٤)

$w_i$	$y_i$	$\hat{y}_i$	$(y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$	$w_i(y_i - \hat{y}_i)^2$
5	10.2	9.97375	0.226246	0.0511874	0.255937
5	10.4	10.6257	-0.225745	0.0509606	0.254803
5	10.6	11.4851	-0.885065	0.801143	4.00571
6	12.5	12.799	-0.299048	0.0894295	0.536577
5	12	13.88567	-1.8857	3.55586	17.7793
4	15	15.1897	-0.189681	0.0359789	0.143916
6	15.17	15.407	-0.237011	0.0561744	0.337046
5	17	16.9283	0.0716764	0.00513751	0.0256876
7	18.71	18.2323	0.477695	0.228192	1.59734
9	19.22	19.1016	0.118373	0.0140122	0.12611
10	20	19.5363	0.463713	0.215029	2.15029
19	20.32	20.6229	-0.302939	0.0917719	1.74367
14	22.07	21.2749	0.79507	0.632137	8.84951
11	22.64	21.9269	0.713079	0.508482	5.5833
9	22.78	22.5789	0.201088	0.0404365	0.363929
14	23.93	23.2309	0.69697	0.485737	6.84232
10	25.5	24.9685	0.530455	0.281383	2.81383
12	25.08	25.8389	-0.758665	0.575878	6.90254
9	26.6	26.4909	0.179143	0.0320922	0.288829
7	28	29.0988	-1.09882	1.20741	8.45185
10	31.67	31.7068	-0.0367846	0.0013531	0.013531
7	32.14	33.0108	-0.870766	0.758234	5.30764

جدول (٢-٤)

source	df	SS	MS	F
regression	1	6164.28	6164.28	1657.24
residual	20	74.3921	3.71960	---
Total	21	6238.67	---	---

ومن جدول تحليل التباين في جدول (٢-٤) وبما أن قيمة  $F$  المحسوبة تزيد عن قيمة  $F$  الجدولية  $F_{0.05}[1,20] = 4.35$  فإننا نرفض فرض العدم. لاختبار جودة التوفيق فإننا نصب مجموع المربعات الصافي أو الخالص والذي يعتبر تقدير ل  $\sigma^2$  لا يعتمد على التوزيع ويتم حسابة من الصيغة التالية :

$$MSPE = \hat{\sigma}^2 = \frac{\sum (n_i - 1) s_i^2}{\sum (n_i - 1)} = \frac{255.2}{167} = 1.528 .$$

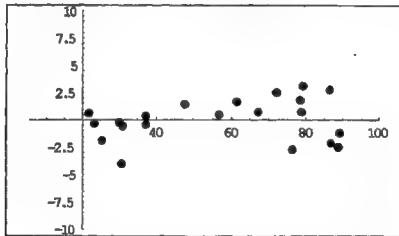
وذلك بـدرجات حرية:

$$\sum (n_i - 1) = 167 .$$

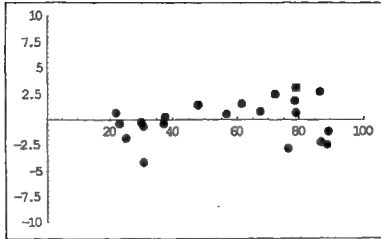
مجموع المربعات لنقص التوفيق هو MSE والمعطى في جدول تحليل التباين في جدول (٤١-٢). وعلى ذلك قيمة F اللازمة لنقص التوفيق تحسب من المعادلة التالية :

$$F = \frac{MSE}{MSPE} = \frac{3.720}{1.528} = 2.43 .$$

وبما أن قيمة F تزيد عن قيمة F الجدولية  $F_{0.01}[20,167] \approx 1.88$  فهذا يعني رفض فرض عدم  $H_0 : \mu_{Y|X_i} = \beta_0 + \beta_1 X_i$  وقبول الفرض البديل  $H_1 : \mu_{Y|X_i} \neq \beta_0 + \beta_1 X_i$ . من البيانات في جدول (٤٠-٢) يتم رسم البواقي المرجحة  $\sqrt{w_i}(y_i - \hat{y}_i)$  مقابل  $\sqrt{w_i} \hat{y}_i$  والمعطاة في شكل (٤٠-٢). أيضا يوضح شكل (٥٥-٢) رسم البواقي المرجحة مقابل  $\sqrt{w_i} X_i$ .



شكل (٤٠-٢)



شكل (٢-٥٥)

يتضح من شكل (٢-٥٤) وشكل (٢-٥٥) أن النقاط تتركز نحو الصفر أي أن طريقة المربعات الصغرى حققت ثبات التباين.

#### طرق أخرى لحساب الأوزان

في كثير من المشاكل فإن الأوزان لا تكون معروفة في البدايه ونحتاج إلى تقديرها بالاعتماد على نتائج المربعات الصغرى العاديه. في الجزء التالي سوف نشرح طريقتين لإيجاد الأوزان  $w_i$ :

##### الطريقة الأولى:

نقسم مشاهدات  $x$  إلى مجاميع متجانسة ويتم حساب تباين كل مجموعة فاضلا عن المتوسط الحسابي لكل مجموعة ثم نستخدم قيم  $s^2$  مقابل قيم  $\bar{x}$  لهذه المجاميع في حساب معادلة الانحدار المقدرة:

$$s_y^2 = b_0 + b_1 \bar{x}, \quad i=1,2,\dots,k. \quad \text{حيث:}$$

والآن لإيجاد التباين لكل مشاهد  $y_i$  نعوض بقيمة  $x_i$  في المعادلة السابقة والتي حصلنا عليها وذلك بإستخدام طريقة المربعات الصغرى العاديه .

الآن الوزن  $w_i$  لكل مشاهدة  $y_i$  هو مقلوب التباين والصوب من المعادله المقدره السابقه .

بعد إيجاد هذه الأوزان نوجد تقدير للمعالم  $\beta_0, \beta_1$  من (٢-١٢) و (٢-١٣).

لاختبار ما إذا كانت طريقة المربعات الصغرى المرجحة قد أدت إلى تجانس التباين نرسم البواقي المرجحة  $\sqrt{w_i}e_i$  مقابل  $\sqrt{w_i}\hat{y}_i$  أو مقابل  $\sqrt{w_i}x_i$  حيث:

$$\sqrt{w_i}e_i = \sqrt{w_i}(y_i - \hat{y}_i).$$

حيث  $\hat{y}_i$  تحسب من المعادلة للنتائج بطريقة المربعات الصغرى المرجحة. إذا كانت البواقي تتركز حول الصفر فهذا يدل على أن طريقة المربعات الصغرى المرجحة صححت مشكلة اختلاف التباين وجعلته متجانساً.

مثال (١٤-٢)

يعطى جدول (٤٢-٢) مقدار المتوسط الشهري لمبيعات مساحه ما (Y) في كل من 30 متجرأ وتكاليف الإعلانات x المقابله السنوية. تهتم الادارة العلاقة بين x , Y والمطلوب إيجاد معادلة الانحدار المقدرة باستخدام طريقة المربعات الصغرى المرجحة إذا كان هناك مخالفه في فرض تجانس التباين .

جدول (٤٢-٢)

$x_i$	$y_i$
3000	81464
3150	72661
3085	72344
5225	90743
5350	98588
6090	96507
8925	126574
9015	114133
8885	115814
8950	123181
9000	131434
11345	140564
12275	151352
12400	146926
12525	130963
12310	144630
13700	147041
15000	179021
15175	166200
14995	180732
15050	178187
15200	185304
15150	155931

16800	172579
16500	188851
17830	192424
19500	203112
19200	192482
19000	218715
19350	214317

الحل

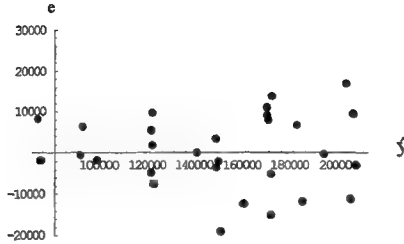
معادلة الانحدار المقدرة بطريقة المربعات الصغرى العادية تم حسابها حيث:

$$\hat{y} = 49443.3838 + 8.0484x$$

البواقي  $(y_i - \hat{y})$  معطاه في جدول (٤٣-٢). رسم البواقي  $e_i$  مقابل  $\hat{y}_i$  معطاه في شكل (٥٦-٢).

جدول (٤٣-٢)

y	$\hat{y}$	e
81464	73588.7	7875.29
72661	74796.	-2134.98
72344	74272.8	-1928.83
90743.	91496.5	-753.501
98588	92502.6	6085.44
96507	98458.4	-1951.41
126574	121276.	5298.26
114133	122000.	-7867.1
115814	120954.	-5139.8
123181	121477.	1704.05
131434	121879.	9554.62
140564	140753.	-188.976
151352	148238.	3113.97
146926	149244.	-2318.08
130963	150250.	-19287.1
144630	148520.	-3889.72
147041	159707.	-12666.1
179021	170170.	8850.96
166200	171579.	-5378.51
180732	170130.	10602.2
178187	170572.	7614.54
185304	171780.	13524.3
155931	171377.	-15446.3
172579	184657.	-12078.2
188851	182243.	6608.3
192424	192947.	-523.132
203112	206388.	-3276.03
192482	203974.	-11491.5
218715	202364.	16351.2
214317	205161.	9136.23



شكل (٥٦-٢)

يتضح من شكل (٥٦-٢) أن هناك مخالفة في فرضي تجانس التباين حيث انتشار البواقي علي شكل القمع المفتوح من الامام وعلي ذلك فإن توفيق المربعات الصغرى العادية يصبح غير ملائم لتصحيح عدم تجانس التباين ولا بد من تقدير المعالم باستخدام طريقة المربعات الصغرى المرجحة .

عند النظر إلى قيم مشاهدات  $x_i$  في جدول (٤٢-٢) نري بأنها مقسمة (نوعاً ما) إلى مجاميع متجانسة ، التباين والمتوسط الحسابي لكل مجموعه معطى في جدول (٤٤-٢).

جدول (٤٤-٢)

$\bar{x}$	$s^2$
3078.3	26794620
5287.5	30722010
8955	52803698
12377.5	77280167
15095	120571040
16650	132388990
19262.5	138856867

ومن قيم  $s^2$  ,  $\bar{x}$  المعطاه في جدول (٢-٤٤) نحسب معادلة الانحدار المقدرة التالية:

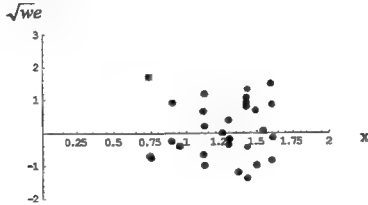
$$s_{\hat{y}}^2 = -7376216.04 + 7819.77\bar{x}.$$

والآن لإيجاد التباين لكل مشاهد  $y_i$  نعوض بقيمة  $x_i$  المقابلة لها في المعادله السابقه ثم نحسب الوزن  $w_i$  لكل مشاهد من معكوس التباين المحسوب من المعادله السابقه للمشاهد  $y_i$ .

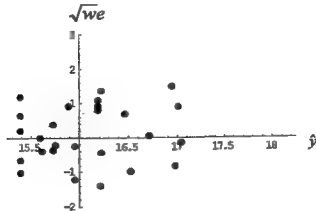
بتطبيق طريقة المربعات الصغرى المرجحه فإن معادلة الانحدار المقدرة سوف تكون:

$$\hat{y} = 50975.5667 + 7.9222x.$$

رسم البواقي المرجحه  $\sqrt{w_i}e_i$  مقابل  $x_i$  ,  $\hat{y}_i$  معطاه في شكل (٢-٥٧) و (٢-٥٨) علي التوالي. من ملاحظه الرسم نرى أن النقاط تتركز حول الصفر وبذلك فإن طريقة المربعات صححت مشكله عدم تجانس التباين.



شكل (٢-٥٧)



شكل (٢-٥٨)



### الطريقة الثانية:

لإيجاد الأوزان  $w_i$  نتبع الخطوات التالية:

١. نحسب معادلة الانحدار المقدرة باستخدام طريقة المربعات العادية ونرسم البواقي مقابل  $\hat{y}_i$  أو  $\hat{x}_i$  وإذا كان انتشار النقاط في رسم البواقي على شكل قمع مفتوح من الأمام أو من الخلف أو على شكل قوسين فهذا يدل على عدم تجانس التباين.

٢. باستخدام طريقة المربعات الصغرى العادية نحسب معادلة الانحدار المقدرة باستخدام القيم المطلقة  $|e_i|$  مقابل قيم  $\hat{x}_i$  أو قيم  $\hat{y}_i$ .

نستخدم معادلة الانحدار المقدرة المحسوبة من الخطوة الثانية في تقدير الأوزان  $w_i$  اللازمة لطريقة المربعات الصغرى المرجحة كما يتضح في المثال التالي :

مثال (٢-١٥)

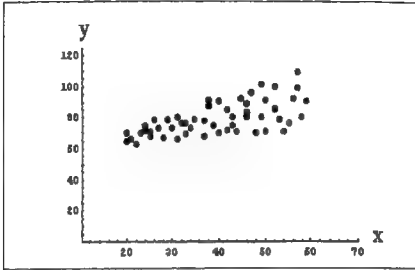
تهتم باحثة صحية بدراسة العلاقة بين ضغط الدم الانبساطي والعمر عند النساء البالغات اللواتي يتمتعن بصحة جيدة وتتراوح أعمارهم بين 20 و 60 عاماً، وقد جمعت بيانات إحصائية عن 54 امرأة والبيانات معطاه في جدول (٢-٤٥).

الحل

يوضح شكل الانتشار المعطى في شكل (٢-٥٩) ان العلاقة بين  $x$  ,  $y$  علاقة خطية.

جدول (۲-۴۵)

x	y	$x^2$	xy
21	66	441	1386
22	63	484	1386
24	75	576	1800
23	70	529	1610
20	65	400	1300
20	70	400	1400
24	72	576	1728
27	73	729	1971
25	71	625	1775
29	79	841	2291
25	68	625	1700
28	67	784	1876
26	79	676	2054
32	76	1024	2432
33	69	1089	2277
31	66	961	2046
34	73	1156	2482
33	76	1089	2508
30	73	900	2190
31	80	961	2480
38	91	1444	3458
37	78	1369	2886
38	87	1444	3306
35	79	1225	2765
37	68	1369	2516
39	75	1521	2925
40	70	1600	2800
42	72	1764	3024
43	80	1849	3440
43	75	1849	3225
44	71	1936	3124
40	90	1600	3600
42	85	1764	3570
46	89	2116	4094
49	101	2401	4949
46	83	2116	3818
46	80	2116	3680
47	96	2209	4512
45	92	2025	4140
49	80	2401	3920
48	70	2304	3360
54	71	2916	3834
52	86	2704	4472
53	79	2809	4187
52	85	2704	4420
50	71	2500	3550
50	91	2500	4550
52	100	2704	5200
55	76	3025	4180
57	99	3249	5643
56	92	3136	5152
59	90	3481	5310
58	80	3364	4640
57	109	3249	6213



شكل (٢-٥٩)

الآن نحسب معادلة الانحدار المقفرة كالتالي:

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$= \frac{173155 - \frac{(2137)(4272)}{54}}{91629 - \frac{(2137)^2}{54}} = \frac{4094.56}{7059.2}$$

$$= 0.580031,$$

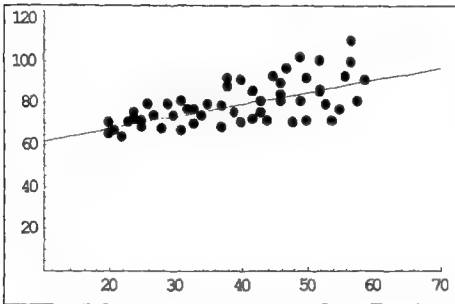
$$b_0 = \bar{y} - b_1 \bar{x}$$

$$= 79.1111 - (0.580031)(39.5741) = 56.1569.$$

معادلة الانحدار المقفرة سوف تكون:

$$\hat{y} = 56.1569 + 0.580031x$$

والممثلة بيانياً في شكل (٢-٦٠) مع شكل الانتشار.



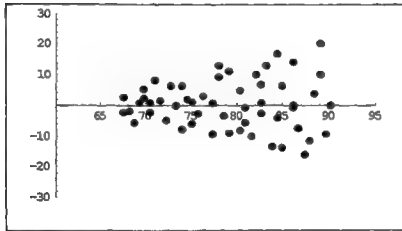
شكل (٦٠-٢)

البواقي  $e_i = y_i - \hat{y}_i$  معطاة في جدول (٤٦-٢) .

جدول (٢-٤٦)

x	y	$\hat{y}$	e
21	66	68.3376	-2.33758
22	63	68.9176	-5.91761
24	75	70.0777	4.92233
23	70	69.4976	0.502362
20	65	67.7575	-2.75755
20	70	67.7575	2.24245
24	72	70.0777	1.92233
27	73	71.8178	1.18224
25	71	70.6577	0.342301
29	79	72.9778	6.02218
25	68	70.6577	-2.6577
28	67	72.3978	-5.39779
26	79	71.2377	7.76227
32	76	74.7179	1.28209
33	69	75.2979	-6.29795
31	66	74.1379	-8.13788
34	73	75.878	-2.87798
33	76	75.2979	0.702054
30	73	73.5579	-0.557853
31	80	74.1379	5.86212
38	91	78.1981	12.8019
37	78	77.6181	0.381931
38	87	78.1981	8.8019
35	79	76.458	2.54199
37	68	77.6181	-9.61807
39	75	78.7781	-3.77813
40	70	79.3582	-9.35816
42	72	80.5182	-8.51822
43	80	81.0983	-1.09825
43	75	81.0983	-6.09825
44	71	81.6783	-10.6783
40	90	79.3582	10.6418
42	85	80.5182	4.48178
46	89	82.8383	6.16165
49	101	84.5784	16.4216
46	83	82.8383	0.161654
46	80	82.8383	-2.83835
47	96	83.4184	12.5816
45	92	82.2583	9.74168
49	80	84.5784	-4.57844
48	70	83.9984	-13.9984
54	71	87.4786	-16.4786
52	86	86.3185	-0.318531
53	79	86.8986	-7.89856
52	85	86.3185	-1.31853
50	71	85.1585	-14.1585
50	91	85.1585	5.84153
52	100	86.3185	13.6815
55	76	88.0586	-12.0586
57	99	89.2187	9.78132
56	92	88.6387	3.36135
59	90	90.3787	-0.378746
58	80	89.7987	-9.79872
57	109	89.2187	19.7813

يوضح رسم البواقي  $e_i$  مقابل  $x_i$  والموضح في شكل (٢-٦١) أن تباين البواقي غير ثابت حيث شكل الانتشار يأخذ شكل القمع المفتوح من الأمام.



شكل (٢-٦١)

الآن نستخدم قيم  $e_i$  و  $x_i$  لأيجاد معادلة الانحدار المقترنة وذلك من البيانات المعطاة في جدول (٢-٤٧).

جول (۴۷-۲)

x	e	x <sup>2</sup>	xe
21	2.33758	441.	49.0891
22	5.91761	484.	130.187
24	4.92233	576	118.136
23	0.502362	529	11.5543
20	2.75755	400	55.1509
20	2.24245	400	44.8491
24	1.92233	576	46.136
27	1.18224	729	31.9205
25	0.342301	625	8.55752
29	6.022178	841	174.643
25	2.657699	625	66.4425
28	5.397792	784	151.138
26	7.76227	676	201.019
32	1.28209	1024	41.0267
33	6.29795	1089	207.032
31	0.13788	961	252.274
34	2.87798	1156	97.0512
33	0.702054	1089	23.1678
30	0.557853	900	16.7356
31	5.86212	961	181.726
38	12.8019	1444	486.472
37	0.381931	1369	14.1315
38	8.8019	1444	334.472
35	2.54199	1225	88.9697
37	9.61807	1369	355.869
39	3.77813	1521	147.347
40	9.35816	1600	374.326
42	8.518222	1764	357.765
43	1.09825	1849	47.2249
43	6.09825	1849	262.225
44	10.67828	1936	469.845
40	10.64183	1600	425.674
42	4.48178	1764	188.235
46	6.16165	2116	283.436
49	16.4216	2401	804.657
46	0.16165	2116	7.43608
46	2.83835	2116	130.564
47	12.5816	2209	591.336
45	9.74168	2025	438.376
49	4.57844	2401	224.343
48	13.9984	2304	671.924
54	16.4786	2916	889.844
52	0.318531	2704	16.5636
53	7.898561	2809	418.624
52	1.31853	2704	68.5636
50	14.1585	2500	707.923
50	5.84153	2500	292.077
52	13.6815	2704	711.436
55	12.0586	3025	663.224
57	9.78132	3249	557.535
56	3.36135	3136	188.235
59	0.378746	3481	22.346
58	9.79872	3364	568.326
57	19.7813	3249	1127.53
2137	339.822	91629	14847.1

$$\Sigma x = 2137, \Sigma |e| = 339.822$$

$$\Sigma x^2 = 91629, \Sigma x|e| = 14847$$

حيث:

$$\begin{aligned} b_1 &= \frac{\Sigma x|e| - \frac{\Sigma x \Sigma |e|}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} \\ &= \frac{14847.1 - \frac{(2137)(339.822)}{54}}{91629 - \frac{(2137)^2}{54}} \\ &= \frac{1398.94}{7059.2} = 0.198172, \\ b_0 &= \frac{\Sigma (|e_i|)}{n} - b_1 \frac{\Sigma x_i}{n} \\ &= 6.29301 - 0.198172(39.5741) \\ &= -1.54998. \end{aligned}$$

وعلى ذلك معادلة الانحدار المقدرة سوف تكون:

$$s = -1.54948 + 0.198172x$$

حيث s تمثل الانحرافات المعيارية.

والآن لأيجاد الانحراف المعياري لكل مشاهدته  $y_i$  نعوض بقيمة  $x_i$  في المعادله السابقه . الوزن  $w_i$  لكل مشاهدته  $y_i$  هو معكوس مربع الانحراف المعياري والمحسوب من المعادله المقدره السابقه والمعطى في جدول (٤٨-٢) .

جدول (٤٨-٢)

s	$1/s^2 = w$
2.61214	0.14657
2.81031	0.126617



3.20666	0.0972512
3.00849	0.110485
2.41397	0.171608
2.41397	0.171608
3.20666	0.0972512
3.80117	0.0692093
3.40483	0.0862599
4.19752	0.0567564
3.40483	0.0862599
3.99935	0.0625204
3.603	0.077032
4.79204	0.0435472
4.99021	0.040157
4.59386	0.0473853
5.18838	0.0371481
4.99021	0.0401571
4.39569	0.0517542
4.59386	0.0473853
5.98107	0.0279539
5.38655	0.0299026
5.7829	0.0279539
5.98107	0.034465
5.38655	0.0299026
5.7829	0.0261896
6.17924	0.0245873
6.37741	0.0217942
6.77376	0.0205728
6.97193	0.0205728
6.97193	0.0194513
7.1701	0.0245837
6.37741	0.0217942
6.77376	0.0174669
7.56645	0.015047
8.16097	0.0174669
7.56645	0.0174669
7.56645	0.0165867
7.76462	0.0165867
7.36828	0.0184191
8.16097	0.0150147
7.96279	0.0157714
9.151839	0.0119395
8.75548	0.0130449
8.95365	0.0124738

8.75548	0.0130449
8.35914	0.0143112
8.35914	0.0143112
8.75548	0.0130449
9.35	0.0114387
9.74634	0.0105273
9.54817	0.0109688
10.1427	0.00972062
9.94452	0.0101119
9.74634	0.0105273

الآن نوجد تقديرات للمعالم  $\beta_0, \beta_1$  كالآتي:

$$b_1 = \frac{\sum x_i y_i w_i - \frac{\sum x_i w_i \sum y_i w_i}{\sum w_i}}{\sum x_i^2 w_i - \frac{(\sum x_i w_i)^2}{\sum w_i}}$$

$$\approx 0.596342,$$

$$b_0 = \frac{\sum y_i w_i}{n} - b_1 \frac{\sum x_i w_i}{n}$$

$$\approx 55.5658.$$

لاختبار ما إذا كانت طريقة المربعات الصغرى المرجحة قد أدت إلى تجانس التباين ومن جنول (٢-٤٩) نرسم البواقي المرجحة  $\sqrt{w_i} e_i$  مقابل  $\sqrt{w_i} \hat{y}_i$  أو مقابل  $\sqrt{w_i} x_i$  والموضحة في شكل (٢-٦٢) وشكل (٢-٦٣) على التوالي.

جول (۴۹-۲)

$w_i$	$\sqrt{w_i}$	$e_i$	$\sqrt{w_i}e_i$	$\sqrt{w_i}\hat{y}_i$	$\sqrt{w_i}x_i$
0.146557	0.382828	-2.08894	-0.799705	26.0663	8.03938
0.126617	0.355832	-5.68528	-2.02301	24.4404	7.82831
0.0972512	0.311851	5.12203	1.59731	21.7915	7.48443
0.110485	0.332393	0.718374	0.238783	23.0287	7.64504
0.171608	0.414256	-2.4926	-1.03257	27.9592	8.28511
0.171608	0.414256	2.5074	1.0387	27.9592	8.28511
0.0972512	0.311851	2.12203	0.661758	21.7915	7.48443
0.0692093	0.263077	1.33301	0.350683	18.8539	7.10307
0.0862599	0.2937	0.525691	0.154396	20.6983	7.34251
0.0567564	0.238236	6.14032	1.46285	17.3578	6.90884
0.0862599	0.2937	-2.47431	-0.726706	20.6983	7.34251
0.0625204	0.250041	-5.26332	-1.31605	18.0688	7.00114
0.077032	0.277546	7.92935	2.20076	19.7254	7.2162
0.0435472	0.20868	1.3513	0.281998	15.5777	6.67775
0.0401571	0.200392	-6.24504	-1.25146	15.0785	6.61295
0.0473853	0.217682	-8.05236	-1.75285	16.1198	6.74813
0.0371481	0.192739	-2.84136	-0.547644	14.6175	6.5531
0.0401571	0.200392	0.754957	0.151288	15.0785	6.61295
0.0517542	0.227495	-0.456628	-0.103742	16.7109	6.82486
0.0473853	0.217682	5.94764	1.29469	16.1198	6.74813
0.0279539	0.167194	12.7732	2.13561	13.0791	6.35338
0.0299026	0.172924	0.36959	0.0639109	13.4241	6.39818
0.0279539	0.167194	8.77325	1.46684	13.0791	6.35338
0.034465	0.185647	2.56227	0.47568	14.1905	6.49766
0.0299026	0.172924	-9.63041	-1.66533	13.4241	6.39818
0.0261896	0.161832	-3.82309	-0.618699	12.7561	6.31145
0.0245873	0.156803	-9.41943	-1.477	12.4532	6.27213
0.0217942	0.147629	-8.61212	-1.27139	11.9006	6.2004
0.0205728	0.143432	-1.20846	-0.173332	11.6479	6.16759
0.0205728	0.143432	-6.20846	-0.890494	11.6479	6.16759
0.0194513	0.139468	-10.8048	-1.50692	11.4092	6.13659
0.0245873	0.156803	10.5806	1.65907	12.4532	6.27213
0.0217942	0.147629	4.38788	0.647776	11.9006	6.2004
0.0174669	0.132162	6.00251	0.793307	10.9691	6.07947
0.0150147	0.122535	16.2135	1.98671	10.3893	6.00419
0.0174669	0.132162	0.00251473	0.000332352	10.9691	6.07947
0.0174669	0.132162	-2.99743	-0.396155	10.9691	6.07947
0.0165867	0.128789	12.4062	1.59776	10.766	6.0531
0.0184191	0.135717	9.59886	1.30273	11.1832	6.10726
0.0150147	0.122535	-4.78651	-0.586513	10.3893	6.00419
0.0157714	0.125584	-14.1902	-1.78296	10.5729	6.02804
0.0119395	0.109268	-16.7682	-1.83223	9.59024	5.90046
0.0130449	0.114214	-0.575536	-0.0657343	9.88815	5.93914
0.0124738	0.111686	-8.17183	-0.912696	9.7359	5.91937
0.0130449	0.114214	-1.57554	-0.179948	9.88815	5.93914
0.0143112	0.11963	-14.3829	-1.72061	10.2143	5.98148
0.0143112	0.11963	5.61715	0.671977	10.2143	5.98148
0.0130449	0.114214	13.4245	1.53326	9.88815	5.93914
0.0114387	0.106952	-12.3646	-1.32241	9.45076	5.88235
0.0105273	0.102603	9.44276	0.968851	9.1888	5.84835
0.0109688	0.104732	3.0391	0.318291	9.31706	5.865
0.00972062	0.0985932	-0.749928	-0.0739377	8.94732	5.817
0.0101119	0.100558	-10.1536	-1.02102	9.06566	5.83236
0.0105273	0.102603	19.4428	1.99488	9.1888	5.84835



شكل (٦٢-٢)



شكل (٦٣-٢)

يتضح من شكل (٦٢-٢) وشكل (٦٤-٢) أن البواقي تتركز حول الصفر وهذا يدل على أن طريقة المربعات الصغرى المرجحة صحت مشكلة اختلاف التباين وجعلته مجانساً. لاختبار فرض العدم  $H_0: \beta_1 = 0$  ضد القرض البديل  $H_1: \beta_1 \neq 0$  نحسب مجموع مربعات البواقي المرجحة من جدول تحليل التباين (٥٠-٢) حيث:

$$\begin{aligned} SSE &= \sum w_i (y_i - \hat{y}_i)^2 \\ &= 76.5135, \end{aligned}$$

مجموع المربعات الكلية سوف تكون:

$$SYY = \sum y_i^2 w_i - \frac{(\sum y_i w_i)^2}{\sum w_i}$$

$$= 159.854.$$

جدول (٥-٢)

Source	df	SS	MS	F
Regression	1	83.3408	83.3408	56.64
Residual	52	76.5135	1.47141	-
Total	53	159.854	-	-

بما أن قيمة F المحسوبة (56.64) تزيد عن قيمة F الجدولي  $F_{0.05}[1,52] \approx 4.08$  فإننا نرفض فرض العدم  $H_0$ .

#### (٦-٢) اختيار التحويلات

##### Choosing a transformations

يتناول هذا البند ثلاث طرق لاختيار التحويلة المناسبة ، الأولى مناسبة لتحويل قيم المتغير التابع  $y_i \geq 0$  ، بينما الثانية مناسبة لتحويل قيم المتغير المستقل أو قيم المتغير التابع أما الأخيرة فمناسبة لتحويل قيم المتغير المستقل x .

#### (١-٦-٢) تحويل قيم المتغير التابع

بفرض إننا نرغب في تحويل قيم y لتصحيح عدم الاعتدال أو عدم ثبات التباين أو عدم خطية دالة الانحدار . العائلة المغيدة من التحويلات هي تحويلة القوى power transformation  $y^\lambda$  حيث  $\lambda$  معلمه مطلوب إيجاد تقدير لها ، فعلى سبيل المثال  $\lambda = \frac{1}{2}$  تعني استخدام تحويلة الجذر التربيعي حيث  $y' = \sqrt{y}$  و  $\lambda = 0$  تعني استخدام التحويلة اللوغاريتمية حيث  $y' = \ln(y)$  . المعيار في تحديد قيمة  $\lambda$  المناسبة لتحويل قيم y هي إيجاد قيمة  $\lambda$  و التي تجعل مجموع مربعات البواقي SSE لانحدار خطي يستند إلى تلك التحويل اصغر ما يمكن .

عندما نحول قيم المتغير التابع فإن  $Y_i'$  تصبح  $f(Y_i)$  وتبعاً لذلك يتغير مقياس المشاهدات المحولة وعلى ذلك لا يمكننا مقارنته  $MSE$  بعد إجراء التحويّلات ونحتاج إلى إجراء بعض التعديلات للتغلب على هذا المشكلة والتي سوف نتناولها في طريقة بوكس - كوكس Box and Cox التالية :

لكل الحالات عندما كل  $y_i \geq 0$  ، قدم بوكس وكوكس ( 1964 ) العائلة التالية من التحويّلات :

$$u_i = \begin{cases} \frac{(y_i^\lambda - 1)}{\lambda \hat{y}^{\lambda-1}} & , \lambda \neq 0 \\ \hat{y} \ln(y_i) & , \lambda = 0 \end{cases}$$

و  $u_i$ ، حيث  $i = 1, 2, \dots, n$  متغير معيارى يترك  $SSE$  غير متأثر بقيمة  $\lambda$  و  $\hat{y} = \ln^{-1} \left[ \left( \frac{1}{n} \right) \sum \ln y_i \right]$  والذي يمثل الوسط الهندسي للبيانات.

تتوافر برامج حاسب آلي لإيجاد قيم  $\lambda$  المناسبة وكبدل يمكن اختبار عدد من قيم  $\lambda$  ( عادة من 10-20 قيمة كافية ) والقيام بالتحويل المقابل لكل منها ثم توفير دالة الانحدار الخطية لقيم  $u_i$  وحساب  $SSE$  لكل توفير ، ومن ثم اختيار القيمة  $\lambda$  التي تجعل  $SSE$  اصغر ما يمكن . ويمكن إجراء بحث أدق في جوار  $\lambda$  عن القيمة التي تجعل مجموع مربعات البواقي اصغر ما يمكن . إلا أن طريقة بوكس - كوكس تستخدم عادة لتكون مرشداً في اختيار تحويله مما لا يترك حاجة للقيم الدقيقة جداً لـ  $\lambda$  . وعلى أي حال يمكن الاستفادة من شكل الانتشار ورسم البواقي وذلك للتحقق من صلاحية التحويل الذي تحدده طريقة بوكس - كوكس .

عادة يتم رسم  $SSE(\lambda)$  مقابل  $\lambda$  وقراءة قيمة  $\lambda$  التي تؤدي إلى جعل  $SSE(\lambda)$  اصغر ما يمكن .  $100\%(1-\alpha)$  فترة ثقة تقريبية للمعلمة  $\lambda$  يمكن الحصول عليها بعد حساب مجموع المربعات والذي يأخذ الصيغة التالية:

$$SS^* = SSE(\lambda) \left( 1 + \frac{t_{\alpha/2}^2(v)}{v} \right)$$

حيث  $v$  درجات الحرية لمجموع المربعات الحرج البواقي ( $v = n - 2$ ) لقيمة  $\lambda$  المختاره وذلك لنموذج الانحدار الخطي البسيط (1-1) وقراءة حدود الثقة للمعلمة  $\lambda$  من الرسم . إذا احتوت فترة الثقة على القيمة  $\lambda = 1$  فهذا يعنى عدم الضرورة إلى تحويله.

مثال (١٦-٢)

البيانات في جدول (٥١-٢) تعطي نتائج عينة عشوائية لمغفرين  $x$  و  $Y$  المطلوب استخدام طريقة بوكس-كوكس لإيجاد تحويله قوى مناسبة لقيم المتغير  $Y$  وحساب SSE من أجل :

$$\lambda = -2, -1, -0.5, 0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 1, 2.$$

ما هو تحويل  $y$  المقترح ؟

جدول (٥١-٢)

رقم المشاهدة	x	y	رقم المشاهدة	x	y
1	679	0.79	27	837	4.20
2	292	0.44	28	1748	4.88
3	1012	0.56	29	1381	3.48
4	493	0.79	30	1428	7.58
5	582	2.70	31	1255	2.63
6	1156	3.64	32	1777	4.99
7	997	4.73	33	370	0.59
8	2189	9.50	34	2316	8.19
9	1097	5.34	35	1130	4.79
10	2078	6.85	36	463	0.51
11	1818	5.84	37	770	1.74
12	1700	5.21	38	724	4.10
13	747	3.25	39	808	3.94
14	2030	4.43	40	790	0.96
15	1643	3.16	41	783	3.29
16	414	0.58	42	406	0.44
17	354	0.17	43	1242	3.24
18	1276	1.88	44	658	2.14
19	745	0.77	45	1746	5.71
20	435	1.39	46	468	0.64
21	540	0.56	47	1114	1.90
22	874	1.56	48	413	0.51
23	1543	5.28	49	1787	8.33
24	1029	0.64	50	3560	14.94
25	710	4.00	51	1495	5.11
26	1434	0.13	52	2221	3.85
			53	1526	3.93

## الحل

قيم  $SSE(\lambda)$  لقيم مختلفة من  $\lambda$  معطاة في جدول (٥٢-٢).

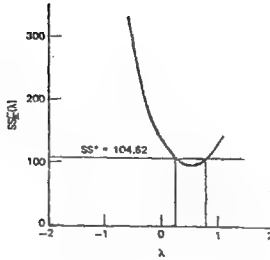
جدول (٥٢-٢)

$\lambda$	$SSE(\lambda)$
-2	34101.0381
-1	986.0423
-0.5	291.5834
0	134.0940
0.125	118.1982
0.25	107.2057
0.375	100.2561
0.5	96.9495
0.625	97.2889
0.75	101.6669
1	126.8660
2	1275.5555

نلاحظ من جدول (٥٢-٢) أن طريقة بوكس-كوكنس تحدد  $\lambda = 0.5$  (تحويله الجذر التربيعي) والقريبه من القيمة المثلى. رسم مجموع مربعات البواقي مقابل  $\lambda$  معطى في شكل (٦٤-٢). عند اختيار  $\lambda = 0.5$  كقيمة مثلى ، فإن 95% فترة ثقة لـ  $\lambda$  يمكن الحصول عليها بحساب مجموع المربعات الحرج الآتي:

$$\begin{aligned}
 SS^* &= SSE(\lambda) \left( 1 + \frac{t_{0.025}^2(v)}{v} \right) \\
 &= 96.9495 \left[ 1 + \frac{(2.095)^2}{51} \right] \\
 &= 104.62.
 \end{aligned}$$





شكل (٦٤-٧)

وعلى ذلك 95% فترة ثقة للمعلمة  $\lambda$  ( من الرسم ) هي :

$$0.26 < \lambda < 0.8$$


حيث 0.26 هو الحد الأدنى للثقة و 0.8 هو الحد الأعلى للثقة . يلاحظ أن فترة الثقة لا تحتوي على القيمة 1 ( التي تعني عدم تحويل البيانات ) ، والذي يعني أن التحويل سوف تكون مفيدة .

(٧-٦-٢) طريقة بيانية لتحويل قيم المتغير التابع أو قيم المتغير المستقل

قدم ( Mosteller and Tukey (1977 طريقة بيانية لاختيار التحويلة المناسبة لقيم المتغير التابع أو قيم المتغير المستقل . تتلخص هذه الطريقة في تقسيم مدى المتغير التابع إلى ثلاثة أقسام والاجتهاد في جعل عدد نقاط المشاهدات في كل قسم متساوية بقدر الإمكان ( تقريبا متساوية ) . في كل فئة من نقاط المشاهدات توجد النقطة ( والتي قد تكون أو لا تكون واحدة من نقاط المشاهدات )

والتي تمثل الفئة تمثيلاً جيداً . لكل فئة فإن الاختيار الجيد لهذه النقطة هي التي احدائياتها الوسيط لقيم  $x$  و  $y$  للنقاط في الفئة . نوجد الميل للخط المستقيم الذي يربط أول نقطتين ( من اليسار إلى اليمين ) والميل الذي يربط الخط المستقيم لأخر نقطتين . فإذا كان الميلين متساويين ، فإن نقاط البيانات لابد أن توصف بخط مستقيم . وإذا لم يتحقق ذلك ، فإن النقطة الوسطي للنقاط الثلاثة سوف تكون تحت (حالة التحدب ) أو فوق ( حالة التقعير ) الخط الذي يربط النقطتين الأخرتين . الآن سوف نجري تحويلة إما لقيم المتغير التابع  $Y$  أو لقيم المتغير المستقل  $x$  وذلك بالاستعانة بتحويلات القوى المعطاة في جدول ( ٢-٥٣ ) .

جدول (٢-٥٣)

⋮		⋮
$-1/y^2$		$x^5$
$-1/y$		$x^4$
$-1/y^{1/2}$		$x^3$
$\log(y)$	مع زيادة التحدب	$x^2$
$y^{1/2}$		$x$
$y$		$x^{1/2}$
$y^2$		$\log(x)$
$y^3$		$-1/x^{1/2}$
$y^4$		$-1/x$
$y^5$	زيادة التقعير	$-1/x^2$
⋮		$-1/x^3$
⋮		⋮

إذا كانت النقاط الثلاثة في شكل محدب فإننا نتحرك إلى أعلى في جدول (٢-٥٣) (فوق السهم  $\rightarrow$ ) ، أما إذا كانوا في شكل مقعر فإننا نتحرك في جدول (٢-٥٣) إلى أسفل (تحت السهم  $\leftarrow$ ) . في كلا الحالتين فإننا نطبق التحويلة للاحداث

المختار ( x أو y ) وذلك للثلاث نقاط . إذا كان الميلين متساويين تقريباً فإننا نتوقف وإذا تغير الشكل من محدب إلى مقعر أو العكس فإننا نتحرك أبعد في الجدول .

مثال (٢-١٧)

للبيانات المعطاة في جدول (٢-٥٤) أختبر هل تحويلة اللوغاريتم لقيم y مناسبة أم لا وذلك باستخدام الطريقة البيانية .

جدول (٢-٥٤)

x	30	50	60	70	75	80	85	90	95	100
y	0.2	1.0	3.0	5.0	8.5	10.0	14.0	20.0	29.0	43.0
log(y)	-0.70	0.00	0.48	0.70	0.93	1.00	1.15	1.30	1.46	1.63
الحل										

بفرض أن النقاط الثلاثة هم :

(50,1) , (77.5,9.25) , (95,29)

الميل بين النقطة الأولى والثانية هو :

$$8.25/27.5 = 3$$

والميل بين النقطة الثانية والثالثة هو :

$$19.75/17.5 = 1.13$$

تحويل قيم المتغير التابع فإننا نصعد في جدول (٢-٥٣) فوق السهم ←

ونأخذ  $y' = \sqrt{y}$  ثم نحصل على الميل لأول نقطتين وآخر نقطتين كالتالي :

$$(3.04 - 1)/27.5 = 0.074 , \quad \frac{2.35}{17.5} = 0.134$$

نلاحظ أن هناك تقارب بين الميلين .

الآن نتحرك إلى أعلى في جدول (٢-٥٣) ونأخذ التحويلة  $y' = \log(y)$

ونحسب الميلين :

$$\frac{0.966 - 0}{27.5} = 0.035 , \quad \frac{0.496}{17.5} = 0.028$$

يلاحظ أن الميلين تقريباً متساويين . ويمكن التحرك إلى أعلى في الجدول

(٢-٥٣) والمحاولة مع  $y' = -1/y^2$  والميل 0.0082 ، 0.025 أى أن الوضع

سوف يكون أسوأ وبذلك فإننا نكتف بتساوي الميل بالتحويلة اللوغاريتمية .

## (٢-٦-٣) تحويل قيم المتغير المستقل

بفرض أن العلاقة بين  $x, Y$  غير خطية ولكن الفرض الخاص بأن حدود الخطأ  $\varepsilon_i$  مستقلة وتتبع توزيعات طبيعية بتباين ثابت تقريبا متحققة . والمطلوب اختيار تحويل تقريبية للمتغير المستقل بحيث أن العلاقة بين  $x, Y$  تكون بسيطة ما أمكن . وصف (Box and Tidwell (1962 طريقة تحليلية لتقدير شكل التحويل لـ  $x$ . وبالرغم من أن هذه الطريقة يمكن استخدامها في حالة الانحدار المتعدد فإننا سوف نطبقها على نموذج الانحدار الخطي البسيط (١-١) .

بفرض أن المتغير التابع  $Y$  يرتبط بـ  $x^\alpha$  ( سوف نضع  $\xi = x^\alpha$  للتسهيل )  
كالتالي :

$$\mu_{Y|x} = E(Y) = f(\xi, \beta_0, \beta_1) = \beta_0 + \beta_1 \xi.$$

$$\xi = \begin{cases} x^\alpha & , \alpha \neq 0 \\ \ln x & , \alpha = 0 \end{cases} \quad \text{حيث:}$$

حيث  $\beta_0$  و  $\beta_1$  معالم مجهولة . بفرض أن  $\alpha_0$  هو القيمة المبدئية ، التي بخمنها القائم على التجربة ، للثابت  $\alpha$ . عادة هذه القيمة المبدئية تكون  $\alpha_0 = 1$  وعلى ذلك  $\xi = x^{\alpha_0} = x$  أو عدم وجود تحويل على الإطلاق تطبق في المحاولة الأولى . باستخدام مفكوك نيولور ومع إهمال الحدود ذات الرتبة العليا فإن :

$$\begin{aligned} \mu_{Y|x} = E(Y) &= f(\xi_0, \beta_0, \beta_1) + (\alpha - \alpha_0) \left\{ \frac{df(\xi, \beta_0, \beta_1)}{d\alpha} \right\}_{\xi=\xi_0, \alpha=\alpha_0} \\ &= \beta_0 + \beta_1 x + (\alpha - 1) \left\{ \frac{df(\xi, \beta_0, \beta_1)}{d\alpha} \right\}_{\xi=\xi_0, \alpha=\alpha_0} . \end{aligned} \quad (١٤-٢)$$

الآن إذا كان الحد بين القوسين في (١٤-٢) معروف فيمكن معاملته كمتغير مستقل مضاف ويكون من الممكن تقدير المعالم  $\beta_0, \beta_1, \alpha$  في (٢-١٤) بطريقة المربعات الصغرى . تقدير  $\alpha$  يمكن اعتباره كتقدير محسن لمعلمة التحويل . الحد بين القوسين في (١٤-٢) يمكن كتابته كالتالي :

$$\left\{ \frac{df(\xi, \beta_0, \beta_1)}{d\alpha} \right\}_{\xi=\xi_0, \alpha=\alpha_0} = \left\{ \frac{df(\xi, \beta_0, \beta_1)}{d\xi} \right\}_{\xi=\xi_0} \left\{ \frac{d\xi}{d\alpha} \right\}_{\alpha=\alpha_0} .$$

وبما أن شكل التحويل معروف  $\xi = x^\alpha$  فإنه  $d\xi/d\alpha = x \ln x$  وعلى ذلك :

$$\left\{ \frac{df(\xi, \beta_0, \beta_1)}{d\xi} \right\}_{\xi=\xi_0} = \frac{d(\beta_0 + \beta_1 x)}{dx} = \beta_1 .$$

المعلمة  $\beta_1$  يمكن تقديرها بتوفيق للنموذج :

$$\hat{y} = b_0 + b_1 x , \quad (١٥-٢)$$

وذلك بطريقة المربعات الصغرى . يمكن تعديل القيمة المبدئية التي فرضناها  $\alpha_0 = 1$  وذلك بتعريف متغير مستقل ثاني هو  $w = x \ln x$  وتقدير المعالم في :

$$\begin{aligned} \mu_{Y|x} &= \beta_0^* + \beta_1^* x + (\alpha - 1)\beta_1 w \\ &= \beta_0^* + \beta_1^* x + \gamma w , \end{aligned}$$

بطريقة المربعات الصغرى والحصول على :

$$\hat{y} = b_0^* + b_1^* x + \hat{\gamma} w \quad (١٦-٢)$$

حيث :

$$\alpha_1 = \frac{\hat{\gamma}}{b_1} + 1 \quad (١٧-٢)$$

وذلك لتقدير جديد لـ  $\alpha$  . ومما يجدر الإشارة إليه أن  $b_1$  نحصل عليها من (١٥-٢) و  $\hat{\gamma}$  من (١٦-٢) . عموماً  $b_1, b_1^*$  سوف يختلفان . الآن نكرر هذه الطريقة باستخدام متغير تابع جديد  $x' = x^{\alpha_1}$  في الحسابات . وقد نوه Box and Tidwell (1962) أن هذه الطريقة عادة ، تميل إلى الالتقاء عند نقطة واحدة بسرعة وعادة المرحلة الأولى تؤدي إلى  $\alpha_1$  كتقدير كافٍ للمعلمة  $\alpha$  . وقد حذر Box and Tidwell (1962) من الإسراف في تدوير الأخطاء فيجب الإبقاء على عدد كافٍ من الأماكن العشرية . مشاكل الالتقاء بنقطة واحدة تظهر عندما يكون الخطأ المعياري  $\sigma$  كبير أو عندما يكون مدى المتغير المستقل صغير جداً بالمقارنة لمتوسطة . هذه الحالة تؤدي إلى بيانات لا تدعم الحاجة لأي تحويل .

مثال (١٨-٢)

استخدم طريقة Box and Tidwell (1952) وذلك لإيجاد تحويل قوى مناسب لقيم المتغير  $x$  وذلك باستخدام البيانات في جدول (٥٥-٢) و التي تمثل نتائج عينة عشوائية لمتغيرين  $Y, x$  .

جدول (٢-٥٥)

رقم الملاحظة	x	y
1	5.00	1.582
2	6.00	1.822
3	3.40	1.057
4	2.70	0.500
5	10.00	2.236
6	9.70	2.386
7	9.55	2.294
8	3.05	0.558
9	8.15	2.166
10	6.20	1.866
11	2.90	0.653
12	6.35	1.930
13	4.60	1.562
14	5.80	1.737
15	7.40	2.088
16	3.60	1.137
17	7.85	2.179
18	8.80	2.112
19	7.00	1.800
20	5.45	1.501
21	9.10	2.303
22	10.20	2.310
23	4.10	1.194
24	3.95	1.144
25	2.45	0.123

الحل

سوف نبدأ بالقيمة البدئية للمعلمة  $\alpha$  وهي  $\alpha_0 = 1$ . نموذج الانحدار المقدر سوف يكون :

$$\hat{y} = 0.1309 + 0.2411x.$$

وعلى ذلك نعرف المتغير  $w = x \ln x$  ونوجد معادلة الانحدار المقدر للنموذج (٢-١٦) والتي سوف تكون :

$$\hat{y} = b_0^* + b_1^* x + \hat{y} w$$

$$= -2.4168 + 1.5344x - 0.4626w . \quad (١٨-٢)$$

المعادلة المقدرة (١٨-٢) تعتبر معادلة انحدار متعدد و سوف نتناول طريقة حسابها في الفصل الثالث . من (١٧-٢) نحسب :

$$\alpha_1 = \frac{\hat{y}}{b_1} + 1 = \frac{-0.4626}{0.2411} + 1 = -0.92 .$$

والتي تعتبر تقدير محسن للمعلمة  $\alpha$  . يلاحظ أن هذا التقدير قريب من القيمة 1- وعلى ذلك فإن تحويله المعكوس لقيم  $x$  تكون مناسبة .

### (٧-٢) وجود مشاهدة واحدة أو قليل من المشاهدات المتطرفة

يطلق مصطلح الخوارج outliers على فئة قليلة من المشاهدات المتطرفة والتي تقع بعيدا عن خط الانحدار، أي تبعد قيمتها بصورة كبيرة عن بقية قيم المشاهدات. وعادة يكون حد الخطأ لها كبير مقارنة ببقية المشاهدات (الطبيعية الأخرى) وأنها قد تؤثر على النموذج الخطي وتقديراته وخصوصا إذا كان حجم العينة صغيرا. يمكن اكتشاف الخوارج برسم البواقي مقابل  $\hat{y}$  ويفضل الاعتماد على البواقي المعيارية. أيضا يمكن استخدام الورق الاحتمالي الطبيعي والذي سوف يكون مفيد في اكتشاف الخوارج. يجب فحص البواقي بعناية وذلك لمعرفة سبب هذا السلوك. في بعض الأحيان تكون الخوارج قيم رديئة bad values تحدث كنتيجة لأحداث غير طبيعية مثل أخطاء في تسجيل المشاهدات أو عطل في جهاز. في هذه الحالة لابد من تصحيح هذه الخوارج (إن أمكن) أو حذفها من فئة المشاهدات. فعلى سبيل المثال فإن واحد أو اثنين من المشاهدات في تجربة كيميائية قد تتأثر بالتلوث الكيميائي أو بعطل في الجهاز. أيضا إذا كانت فئة من المشاهدات تمثل أسعار بيع منازل فقد تشتمل هذه الفئة على منزل أو منزلين تم بيعهما بأسعار لا تعكس الحقيقة لأي سبب من الأسباب. في بعض الأحيان نجد أن الخوارج غير طبيعية ولكن مفيدة وإن إزالة تلك النقاط لتحسين التوفيق للمعادلة يكون من الخطورة، حيث نجد أن الفئة المتطرفة أكثر أهمية من باقي الفئة من المشاهدات. كثير من الاختبارات الإحصائية تستخدم في اكتشاف ورفض الخوارج والمقترح، على سبيل المثال من قبل

, Ellenberg (1976) , Anscombe (1960)

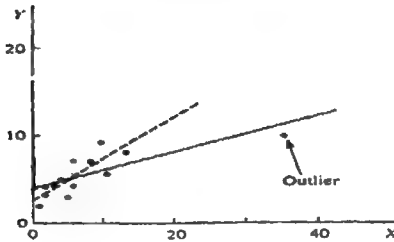
Anscombe and Tukey (1963). وهناك اختبار تقريبي قدم من قبل Stefansky (1971-1972) لتحديد الخوارج يعتمد على القيمة العظمى التالية:

$$\frac{|e_j|}{\sqrt{\frac{\sum_{j=0}^n e_j^2}{n}}},$$

ومن السهل تطبيقه.

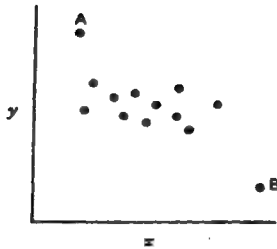
للتحقق من تأثير البواقي نقوم بتحليل البيانات كاملة ومن ثم تحليل البيانات باستثناء الخواارج ومقارنة  $MSE$  و  $F$  و  $R^2$  في كلتا الحالتين. يوضح شكل (٦٥-٢) الانتشار لعدد 13 حالة حيث توجد مشاهدة واحدة متطرفة عن بقية قيم المشاهدات. والخط المستقيم على الرسم يمثل معادلة الانحدار المقدرة لكل فئة المشاهدات. الخط المنقط يمثل معادلة الانحدار المقدرة بعد حذف المشاهدات المتطرفة. من الواضح أن خط التوفيق تأثر كثيراً بوجود القيمة المتطرفة.

شكل (٦٥-٢)



إذا كانت هذه المشاهدات ناتجة من خطأ في تسجيل البيانات فلا بد من تصحيحها أو حذفها من فئة المشاهدات قبل توفيق خط الانحدار. في بعض الحالات تكون القيمة المتطرفة ليس لها تأثير ففي شكل (٦٦-٢) فإن القومتين المتطرفتين تقعان على توفيق خط الانحدار. لاحظ أن خط التوفيق هو نفسه سواء أزيلت القيمة المتطرفة أولاً.





شكل (٦٦-٢)

مثال (١٩-٢)

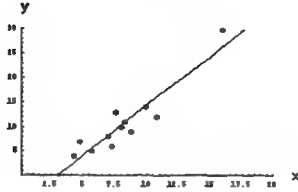
لازواج القياسات في جدول (٥٦-٢) أوجد شكل الانتشار وناقش وجود قسيم منظرية.

جدول (٥٦-٢)

x	y
10	8.04
8	6.95
13	7.58
9	8.81
11	8.33
14	9.96
6	7.24
4	4.24
12	10.84
7	4.82
5	5.68
16	30

الحل

شكل الانتشار للبيانات المعطاة في جدول (٥٦-٢) موضح في شكل (٦٧-٢).



شكل (٦٧-٢)

يتضح من شكل (٦٧-٢) أن هناك قيمة متطرفة بعيدة عن خط الانحدار. ومعادلة خط الانحدار المقدرة هي  $\hat{y} = -0.997531 + 1.33284x$ . جدول تحليل التباين معطى في جدول (٥٧-٢).

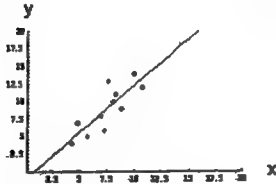
جدول (٥٧-٢)

Source	df	SS	MS	F
الانحدار	1	444.643	444.643	63.8787
الخطأ	10	69.6074	6.9674	--
الكلى	11	514.25	--	--

سوف نوجد معادلة الانحدار المقدرة بنون القيمة المتطرفة  $x=16$ ,  $y=30$  حيث معادلة الانحدار المقدرة هي:

$$\hat{y} = -3.94649 + 2.03388x.$$

شكل الانتشار موضح في شكل (٦٨-٢).



شكل (٦٨-٢)

جدول تحليل التباين في هذه الحالة معطى في جدول (٥٨-٢)

جدول (٥٨-٢)

Source	df	SS	MS	F
الانحدار	1	73.3197	73.3197	17.9899
الخطأ	9	36.6803	4.07559	---
الكلي	10	110	---	---

وبمقارنة شكل الانتشار نجد أن القيمة المتطرفة لم تثر كثيرا على خط الانحدار. كما أن  $MSE=4.07559$  بدون القيمة المتطرفة و  $MSE=6.9074$  بالقيمة المتطرفة والاختلاف ليس كبيرا. كما أن  $R^2 = 0.8646$  بالقيمة المتطرفة و  $R^2 = 0.6665$  بدون القيمة المتطرفة.

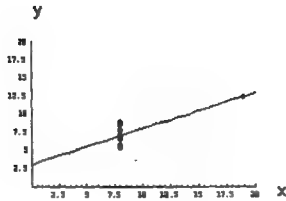
مثال (٢٠-٢)

لأزواج المشاهدات في جدول (٥٩-٢) ارسم شكل الانتشار وناقش وجود الخوارج.

جدول (٥٩-٢)

x	y
8	6.58
8	5.76
8	7.71
8	8.84
8	8.47
8	7.04
8	5.25
19	12.5
8	5.56
8	7.91
8	6.89

شكل الانتشار للملاحظات المعطاة في جدول (٥٩-٢) موضح في شكل (٦٩-٢).



شكل (٦٩-٢)

في هذا الشكل لا يوجد معلومات كافية وذلك لاتخاذ قرار بشأن النموذج الموفق. أن الميل في وجود القيمة المتطرفة سوف يكون كبير جدا. بحذف القيمة الثامنة لن تستطيع التقدير هنا ولا بد من أن نرتاب من إجراء تحليل يعتمد بقل على قيمة واحده.

## الفصل الثالث

### الإنحدار الخطي المتعدد

### Multiple Linear Regression

مقدمة	(١-٣)
تقدير المعالم	(٢-٣)
تقدير المعالم باستخدام المصفوفات	(٣-٣)
الإنحدار البسيط في صيغة مصفوفة	(٤-٣)
فروض جاوس - ماركوف	(٥-٣)
خواص مقدرات المربعات الصغرى	(٦-٣)
خواص البواقي	(٧-٣)
صيغة أخرى للحصول على تقديرات المربعات الصغرى لمعالم نموذج الإنحدار الخطي المتعدد	(٨-٣)
تقدير $\sigma^2$	(٩-٣)
فترات ثقة في الإنحدار المتعدد	(١٠-٣)
فترات ثقة لمعاملات الإنحدار	(١١-٣)
فترة ثقة لمتوسط الاستجابة	(١٢-٣)
فترة ثقة لملاحظة مستقبلية	(١٣-٣)
فترة ثقة لدالة خطية لعدة معاملات إنحدار	(١٤-٣)
تقديرات أو تنبؤات خارج مجال النموذج	(١٥-٣)
إختبارات الفروض	(١٦-٣)
إختبار يخص جميع معاملات الإنحدار الجزئية	(١٧-٣)
معامل التحديد المتعدد	(١٨-٣)
إختبارات تخص كل معامل الإنحدار	(١٩-٣)
طريقة مجاميع المربعات الإضافية	(٢٠-٣)
إختبار فرضية حول أهمية تعاقب المتغيرات	(٢١-٣)
الحالة الخاصة لأعمدة متعامدة في المصفوفة $X'$	(٢٢-٣)
إختبار الفرض الخطي $TB = 0$	(٢٣-٣)
معاملات الإنحدار القياسية	(٢٤-٣)
معامل الارتباط الجزئي من الرتبة الأولى	(٢٥-٣)
معامل الارتباط الجزئي من الرتبة الثانية	(٢٦-٣)

## (١-٣) مقدمة

في الغالب تكون العلاقات الفعلية سواء الاقتصادية أو الاجتماعية أو السياسية معقدة يمثل فيها متغير واحد تابع وعدد من المتغيرات المستقلة. ومن الأمثلة العديدة على ذلك في مجال الاقتصاد نجد أن الكمية المستهلكة من سلعة ما تتأثر بسعر السلعة ذاتها علاوة على أسعار السلع البديلة وأيضاً بالإضافة إلى ذوق المستهلك. كذلك كمية الإنتاج تتأثر بالعمل ورأس المال والموارد الوسيطة وغيرها من عناصر العملية الإنتاجية. وفي مجال التأمين يتوقف القسط التأميني على عمر المؤمن ودخله وقيمة الوثيقة وطول فترات التأمين.

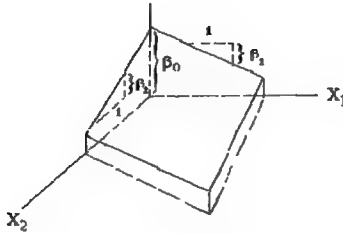
نموذج الانحدار الذي يحتوي على أكثر من متغير مستقل يسمى نموذج الانحدار المتعدد. في هذا الفصل سوف نتناول التوفيق والتحليل لتلك النماذج . النتائج تعتبر تعميم لما تناولناه لنموذج الانحدار الخطي البسيط .

في حالة  $k$  من المتغيرات المستقلة  $x_1, x_2, \dots, x_k$  فإن متوسط المتغير  $Y | x_1, x_2, \dots, x_k$  يعطى بمعادلة إنحدار المجتمع التالية:

$$\mu_{Y|x_1, x_2, \dots, x_k} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (١-٣)$$

حيث  $\beta_0$  يمثل القيمة المتوسطة للمتغير  $Y$  عندما تكون جميع قيم المتغيرات المستقلة مساوية للصفر ويصعب تفسير  $\beta_0$  إذا كانت قيمته سالبة وقيم المتغير التابع الفعلية موجبة. بينما المعامل  $\beta_i$  يمثل التغير في القيمة المتوسطة للمتغير التابع الناتج عن زيادة المتغير المستقل  $x_i$  بمقدار الواحد بافتراض ثبات قيم المتغيرات الأخرى  $x_j$  و  $i \neq j$ . معاملات الانحدار  $\beta_i$  (حيث  $i = 1, 2, \dots, k$ ) تسمى في بعض الأحيان معاملات الانحدار الجزئية.

إن التمثيل البياني للنموذج (١-٣) هو سطح ذو أبعاد  $k+1$  حيث  $k$  تمثل عدد المتغيرات المستقلة. ففي حالة وجود متغيرين مستقلين ( $k = 2$ ) فإن السطح الملامك للبيانات هو سطح ذو ثلاثة أبعاد كما هو موضح في شكل (١-٣).



شكل (١-٣)

ويمكن الحصول على الإستجابة المقدرة من معادلة الإنحدار المقدرة التالية:

$$\hat{y} = b_0 + b_1x_1 + \dots + b_kx_k ,$$

حيث كل معلمة  $\beta_i$  تقدر بواسطة  $b_i$  من بيانات العينة وذلك باستخدام طريقة المربعات الصغرى.

نفس أسلوب المربعات الصغرى يمكن تطبيقه في تقدير معاملات الإنحدار عندما يشتمل النموذج الخطي على أس أو حواصل ضرب للمتغيرات المستقلة. على سبيل المثال ، عندما  $k=1$  ، وعندما يشعر القارئ على التجربة أن المتوسطات  $\mu_{y|x}$  لاتقع على خط مستقيم ولكن يمكن تمثيلها بنموذج إنحدار لكثيرات الحدود التالي:

$$\mu_{y|x} = \beta_0 + \beta_1x + \beta_2x^2 + \dots + \beta_kx^k \quad (٢-٣)$$

حيث :

$$x_1 = x, x_2 = x^2, \dots, x_k = x^k$$

وعلى ذلك (٢-٣) يمكن كتابتها كالتالي:

$$\mu_{y|x} = \beta_0 + \beta_1x_1 + \dots + \beta_kx_k$$

والذي يعتبر نموذج إنحدار خطي متعدد بعدد  $k$  من المتغيرات المستقلة.

كثيراً ما يحدث لبس عندما نتكلم عن نموذج كثيرات الحدود كنموذج إحدار خطي. عادة ما يشير الإحصائيين إلى النموذج الخطي كنموذج إحدار خطي في المعالم وذلك بصرف النظر عن الكيفية التي تظهر فيها المتغيرات المستقلة في النموذج. نماذج كثيرات الحدود سوف نتناولها بالتفصيل في الفصل السادس. النماذج التي تشمل على تأثيرات تفاعل أيضاً يمكن تمثيلها بنماذج الإحدار الخطي المتعدد. على سبيل المثال بفرض النموذج التالي:

$$\mu_{Y|X} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 \quad (٣-٣)$$

$$x_3 = x_1 x_2, \quad \beta_3 = \beta_{12} \quad \text{وبوضع:}$$

فإن (٣-٣) يمكن كتابته على الشكل التالي:

$$\mu_{Y|X} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

### (٢-٣) تقدير المعالم

سوف نستخدم طريقة المربعات الصغرى لتقدير المعالم (معاملات الإحدار) في (١-٣)، وذلك بإستخدام نقاط العينة:

$$\{(x_{1j}, x_{2j}, \dots, x_{kj}, y_j) \mid j = 1, 2, \dots, n, n > k\}$$

حيث  $y_j$  الإستجابة المشاهدة للقيم  $(x_{1j}, x_{2j}, \dots, x_{kj})$  للمتغيرات المستقلة

$x_1, x_2, \dots, x_k$ . كل مشاهدة  $(x_{1j}, x_{2j}, \dots, x_{kj})$  تحقق المعادلة:

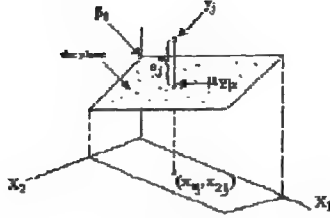
$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \dots + \beta_k x_{kj} + e_j^* \quad (٤-٣)$$

أو المعادلة:

$$y_j = b_0 + b_1 x_{1j} + b_2 x_{2j} + \dots + b_k x_{kj} + e_j \quad (٥-٣)$$

حيث  $e_j^*$  ويمثلان خطأ عشوائي ويبقى على التوالي والمرتبطة بالإستجابة  $y_j$ . المعادلة (٤-٣) موضحة بيانياً في شكل (٢-٣) عندما  $k=2$ .





شكل (٢-٣)

سوف نكتب نموذج العينة (٥-٣) كالتالي:

$$y_j = b_0 + \sum_{i=1}^k b_i x_{ij} + e_j.$$

مجموع المربعات للبقايا هو:

$$SSE = \sum_{j=1}^n e_j^2 = \sum_{j=1}^n \left( y_j - b_0 - \sum_{i=1}^k b_i x_{ij} \right)^2.$$

بإستخدام مفهوم المربعات الصغرى سوف نحصل على التقديرات  $b_0, b_1, \dots, b_k$  التي تجعل SSE أقل ما يمكن ويتم ذلك بإجراء التفاضل الجزئي لـ SSE بالنسبة لكل من  $b_0, b_1, \dots, b_k$  والمساواة بالصفر لنحصل على فئة من  $p = k + 1$  من المعادلات الطبيعية:

$$\begin{aligned} nb_0 + b_1 \sum_{j=1}^n x_{1j} + b_2 \sum_{j=1}^n x_{2j} + \dots + b_k \sum_{j=1}^n x_{kj} &= \sum_{j=1}^n y_j \\ b_0 \sum_{j=1}^n x_{1j} + b_1 \sum_{j=1}^n x_{1j}^2 + b_2 \sum_{j=1}^n x_{1j}x_{2j} + \dots + b_k \sum_{j=1}^n x_{1j}x_{kj} &= \sum_{j=1}^n x_{1j}y_j \\ \vdots & \quad \quad \quad \vdots \\ b_0 \sum_{j=1}^n x_{kj} + b_1 \sum_{j=1}^n x_{kj}x_{1j} + b_2 \sum_{j=1}^n x_{kj}x_{2j} + \dots + b_k \sum_{j=1}^n x_{kj}^2 &= \sum_{j=1}^n x_{kj}y_j \end{aligned}$$

(٦-٣)

تلك المعادلات يمكن حلها لإيجاد  $b_0, b_1, b_2, \dots, b_k$  وذلك بأي طريقة مناسبة لحل نظام من المعادلات. وهناك طريقة سهلة لكتابة المعادلات الطبيعية لأي نموذج والتي سوف نوضحها عندما يكون لدينا 5 متغيرات مستقلة وهي كالتالي:  
للحصول على المعادلة الأولى للمربعات الصغرى نفذ عملية التجميع على حدود طرفي المعادلة التالية:

$$y_j = b_0 + b_1 x_{1j} + b_2 x_{2j} + b_3 x_{3j} + b_4 x_{4j} + b_5 x_{5j}. \quad (٧-٣)$$

اضرب كل حد من حدود المعادلة (٧-٣) في  $x_{1j}$  ثم أجر عملية التجميع للطرفين لنحصل على معادلة المربعات الصغرى الثانية. اضرب حدود المعادلة (٧-٣) في  $x_{2j}$  ثم أجر عملية التجميع على الطرفين لنحصل على المعادلة الثالثة. نكرر الخطوات السابقة نفسها للحصول على مجموعة معادلات المربعات الصغرى وذلك بالضرب في  $x_{3j}$  ثم التجميع ثم نضرب في  $x_{4j}$  ونجري عملية التجميع ثم نضرب في  $x_{5j}$  والتجميع ، عندئذ نحصل على ست معادلات للمربعات الصغرى.

#### مثال (١-٣)

يتأثر الاستهلاك السنوي للطعام (مئات الدولارات) على كل من الدخل السنوي للأسرة  $x_1$  بمئات الدولارات وحجم الأسرة (عدد الأفراد في الأسرة الواحدة)  $x_2$ . أوجد معادلة الإنحدار الخطي المقدرة وذلك بفرض توفر البيانات المعطاة في جدول (١-٣).

جدول (١-٣)

$x_1$	$x_2$	$y$
8	6	22
10	7	23
7	5	18
2	2	9
4	3	14
6	4	20
7	4	21
6	3	18
4	3	16
6	3	19

## الحل

من البيانات في جدول (١-٣) فإن:

$$\begin{aligned} n=10 \quad & \text{و} \quad \Sigma x_{1j}=60 \quad \text{و} \quad \Sigma x_{2j}=40 \\ \Sigma x_{1j}^2=406 \quad & \text{و} \quad \Sigma x_{1j}x_{2j}=269 \quad \text{و} \quad \Sigma x_{2j}^2=182 \\ \Sigma y_j=180 \quad & \text{و} \quad \Sigma x_{1j}y_j=1159 \quad \text{و} \quad \Sigma x_{2j}y_j=766 \quad \text{و} \quad \Sigma y_j^2=3396. \end{aligned}$$

وبالتعويض بالقيم السابقة في المعادلات الطبيعية الثلاثة نحصل على :

$$10b_0 + 60b_1 + 40b_2 = 180$$

$$60b_0 + 406b_1 + 269b_2 = 1159$$

$$40b_0 + 269b_1 + 182b_2 = 766.$$

لهذه الفئة من المعادلات نحصل على التقديرات التالية والوحيدة :

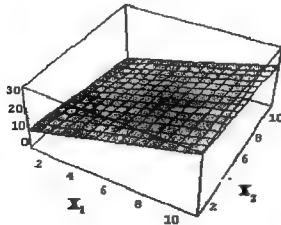
$$b_0 = 7.918, \quad b_1 = 2.363, \quad b_2 = -1.024$$

وعلى ذلك نموذج الإنحدار الخطي المقدر يمكن كتابته على الشكل:

$$\hat{y} = 7.918 + 2.363x_1 - 1.024x_2. \quad (٨-٣)$$

وهذا يعني أنه في كل زيادة مائة دولار في الدخل السنوي للأسرة فإن كمية الإستهلاك السنوي للطعام (بمئات الدولارات) تزيد بمقدار 2.363 وذلك عند ثبات  $x_2$  ، أيضاً أنه في كل زيادة فرد في الأسرة فإن كمية الإستهلاك السنوي للطعام تنقص بمقدار 1.024 وذلك عند ثبات  $x_1$  . التمثيل البياني للمعادلة (٨-٣) موضح في شكل (٣-٣) .

شكل (٣-٣)



### (٣-٣) تقدير المعالم باستخدام المصفوفات

عند توفيق نموذج الانحدار الخطي المتعدد وخصوصاً عندما يزيد عدد المتغيرات المستقلة عن اثنين ، فإن معلومتنا في نظرية المصفوفات يمكن أن تسهل العمليات الحسابية. بفرض أن القائم على التجربة لديه  $k$  من المتغيرات المستقلة  $x_1, x_2, \dots, x_k$  و  $n$  من المشاهدات  $y_1, y_2, \dots, y_n$  وكل مشاهدة يمكن التعبير عنها كالتالي:

$$y_j = b_0 + b_1x_{1j} + b_2x_{2j} + \dots + b_kx_{kj} + e_j, \quad j = 1, 2, \dots, n. \quad (٩-٣)$$

هذا النموذج يمثل  $n$  من المعادلات. باستخدام رموز المصفوفات يمكن كتابة النموذج (٩-٣) على الشكل التالي:

$$y = Xb + e, \quad (١٠-٣)$$

حيث:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix},$$

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}.$$

عوماً،  $y$  متجه من الدرجة  $n \times 1$  من المشاهدات  $X$  مصفوفة من مستويات المتغيرات المستقلة من الدرجة  $n \times p$  ( $p = k + 1$ ) حيث  $p$  عدد المعالم في النموذج و  $b$  متجه من الدرجة  $p \times 1$  من معاملات الانحدار و  $e$  متجه من الدرجة  $n \times 1$  من البواقي والمطلوب إيجاد المتجه من تقديرات المربعات الصغرى  $b$  ، والذي يؤدي إلى تصغير :

$$SSE = \sum e_i^2 = e'e = (y - Xb)'(y - Xb).$$

يمكن التعبير عن SSE كالتالي:

$$\begin{aligned} e'e &= y'y - b'X'y - y'Xb + b'X'Xb \\ &= y'y - 2b'X'y + b'X'Xb. \end{aligned}$$

وذلك لأن  $b'X'y$  مصفوفة بدرجة  $(1 \times 1)$  أو ثابت ومنقولها هو  $y'Xb = (b'X'y)'$  والذي نفسه ثابت. تقديرات المربعات الصغرى هي التي تحقق:

$$\frac{\partial SSE}{\partial b} = -2X'y + 2X'Xb = 0,$$

والذي يبسط إلى:

$$X'Xb = X'y. \quad (١١-٣)$$

المعادلات في (١١-٣) تمثل معادلات المربعات الصغرى. لحل المعادلات الطبيعية نضرب طرفي المعادلة (١١-٣) بالمعكوس للمصفوفة  $X'X$ . وعلى ذلك تقديرات المربعات الصغرى للمتجه  $\beta$  هو:

$$b = (X'X)^{-1}X'y, \quad (١٢-٣)$$

تحت شرط أن المصفوفة  $(X'X)^{-1}$  موجودة حتى يمكن الحصول على حل وحيد. إن المصفوفة  $(X'X)^{-1}$  دائماً تكون موجودة في حالة عدم وجود أي عمود في المصفوفة  $X$  يمكن الحصول عليه كتركيبية خطية من الأعمدة الباقية. وبصورة أخرى المصفوفة  $X'X$  يكون لها محدد لا يساوي صفر. وفي هذه الحالة يقال أن صيغة المصفوفة للمعادلات الطبيعية في (١١-٣) هي نفسها في (٦-٣). بكتابة (١١-٣) بالتفصيل نحصل على:

$$\begin{bmatrix} n & \sum_{j=1}^n x_{1j} & \dots & \sum_{j=1}^n x_{kj} \\ \sum_{j=1}^n x_{1j} & \sum_{j=1}^n x_{1j}^2 & \dots & \sum_{j=1}^n x_{1j}x_{kj} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^n x_{kj} & \sum_{j=1}^n x_{kj}x_{1j} & \dots & \sum_{j=1}^n x_{kj}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n y_j \\ \sum_{j=1}^n x_{1j}y_j \\ \vdots \\ \sum_{j=1}^n x_{kj}y_j \end{bmatrix}$$

المصفوفة  $X'X$  تمثل مصفوفة من الدرجة  $(p \times p)$  ومتماثلة والمصفوفة  $X'y$  متجه من الدرجة  $(p \times 1)$ .

نموذج الانحدار المقدر المقابل لمستويات المتغيرات المستقلة  $\underline{x}'_j = [1, x_{1j}, x_{2j}, \dots, x_{kj}]$  هو:

$$\hat{y} = \underline{x}'_j b.$$

$$= b_0 + \sum_{i=1}^k b_i x_{ij}.$$

وكما في حالة الإنحدار البسيط فإن الفرق بين القيمة المشاهدة  $y_j$  والقيمة المقابلة المقدرة  $\hat{y}_j$  هو الباقي  $e_j$  حيث  $e_j = y_j - \hat{y}_j$ . البواقي التي عددها  $n$  يمكن كتابتها في صيغة مصفوفة كالتالي:

$$e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}, \quad \hat{y} = \begin{pmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{pmatrix} = Xb = X(X'X)^{-1}X'y = Hy, \quad (١٣-٣)$$

حيث:

$$H = X(X'X)^{-1}X'.$$

المصفوفة  $H$  من الدرجة  $n \times n$ . وتعرف المصفوفة  $H$  بمصفوفة القبعة  $\hat{y}$  matrix لأنها تحول قيم المتغير التابع المشاهد  $y$  إلى القيم المقدرة  $\hat{y}$  التي تظهر فوقها علامة القبعة  $\hat{\phantom{y}}$ ، حيث تعرف المصفوفة  $H$  بمصفوفة أيديموتنت idempotent matrix (الجامدة) التي لها الخصائص التالية:

$$H'H = HH = H \quad \text{و} \quad H' = H.$$

ليكن:  $M = I - H$  حيث المصفوفة  $M$  أيضا مصفوفة أيديموتنت فإن:

$$MX = (I - H)X = X - X(X'X)^{-1}X'X = X - X = 0, \quad (١٤-٣)$$

(١٤-٣) الإنحدار البسيط في صيغة مصفوفة

عند وجود متغير مستقل واحد في نموذج الإنحدار الخطي البسيط فإن:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \text{و} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

وعلى ذلك:

$$(X'X) = \begin{bmatrix} n & \sum x_j \\ \sum x_j & \sum x_j^2 \end{bmatrix} \quad \text{و} \quad X'y = \begin{bmatrix} \sum y_j \\ \sum x_j y_j \end{bmatrix}$$

المصفوفة  $(X'X)^{-1}$  يمكن الحصول عليها كالتالي:

$$(X'X)^{-1} = \frac{1}{SXX} \begin{pmatrix} \sum x_j^2 / n & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

وعلى ذلك:

$$\begin{aligned} b &= \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = (X'X)^{-1} X'y \\ &= \frac{1}{SXX} \begin{pmatrix} \sum x_j^2 / n & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} \sum y_j \\ \sum x_j y_j \end{pmatrix} \\ &= \begin{pmatrix} \bar{y} - b_1 \bar{x} \\ \frac{SXY}{SXX} \end{pmatrix} . \end{aligned}$$

تبين  $B_0$  يمكن الحصول عليه كالتالي:

$$\text{Var}(B_0) = \sigma^2 c_{11}$$

حيث  $c_{11}$  هو العنصر رقم 1 على القطر الرئيسي للمصفوفة  $(X'X)^{-1}$ . أي أن :

$$\begin{aligned} \text{Var}(B_0) &= \sigma^2 \left[ \sum x_j^2 / (n) SXX \right] \\ &= \sigma^2 [1/n + \bar{x}^2 / SXX], \end{aligned}$$

وهو نفسه الذي حصلنا عليه في الفصل الأول. وبنفس الشكل تبين  $B_1$  هو:

$$\text{Var}(B_1) = \sigma^2 c_{22} = \frac{\sigma^2}{SXX} .$$

حيث  $c_{22}$  هو العنصر رقم 2 على القطر الرئيسي للمصفوفة  $(X'X)^{-1}$ .

## (٣-٥) فروض جاوس - ماركوف

نموذج الانحدار الخطي المتعدد باستخدام المصفوفات في حالة  $k$  من المتغيرات المستقلة يأخذ الصيغة التالية:

$$Y = X\beta + \varepsilon \quad (٣-١٥)$$

حيث :

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}, \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}_{n \times p},$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{p \times 1}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1}$$

حيث  $p = k+1$ .

لتقدير  $\beta$  فلا بد من تحقق الفروض التالية، والميماة فروض جاوس - ماركوف ، حيث أن حدود الخطأ  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  غير مرتبطة وكل حد له متوسط صفر وتباين  $\sigma^2$  وذلك كما هو الحال في نموذج الانحدار البسيط والتي تناولناه في الفصل الأول ، أي أن:

$$E(\varepsilon_j) = 0 \quad (٣-١٦)$$

$$E(\varepsilon_j^2) = \sigma^2 \quad (٣-١٧)$$

$$E(\varepsilon_j \varepsilon_l) = 0, \quad j \neq l. \quad (٣-١٨)$$

سوف نعبر عن هذه الفروض في صيغة مصفوفات كالتالي:

$$E(\varepsilon) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

حيث  $\varepsilon$  متجه من الأصفار من الدرجة  $(n \times 1)$ . أيضاً :



$$E(\varepsilon \varepsilon') = E \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \cdots & \varepsilon_n \end{bmatrix} ,$$

$$= E \begin{bmatrix} \varepsilon_1^2 & \varepsilon_1 \varepsilon_2 & \cdots & \cdots & \varepsilon_1 \varepsilon_n \\ \varepsilon_1 \varepsilon_2 & \varepsilon_2^2 & \cdots & \cdots & \varepsilon_2 \varepsilon_n \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ \varepsilon_n \varepsilon_1 & \varepsilon_n \varepsilon_2 & \cdots & \cdots & \varepsilon_n^2 \end{bmatrix} = \sigma^2 I_n$$

(١٩-٣)

حيث  $I_n$  مصفوفة الوحدة من الدرجة  $n \times n$ .

تبعا لفروض جاوس - ماركوف فإن :

$$E(Y) = X\beta, \quad (٢٠-٣)$$

$$\text{Cov}(Y) = E(Y - X\beta)(Y - X\beta)' \quad (٢١-٣)$$

$$= E(\varepsilon \varepsilon') = \sigma^2 I.$$

### (٦-٣) خواص مقدرات المربعات الصغرى

المتوسط والتباين للمقدرات  $B_0, B_1, \dots, B_k$  يمكن الحصول عليها تحت شروط تحقق فروض جاوس - ماركوف .

من (١٢-٣) فإن :

$$E(B) = E\{(X'X)^{-1} X'Y\} = (X'X)^{-1} X'X\beta = \beta ,$$

وذلك لأن  $(X'X)^{-1} XX' = I$  ،  $E(\varepsilon) = 0$  ، أي أن  $B$  مقدر غير متحيز لـ  $\beta$  .

مصفوفة التباين يمكن للتعبير عنها كالتالي:

$$\text{Cov}(B) = E\{ [B - E(B)] [B - E(B)]' \} .$$

أيضاً تسمى  $\text{Cov}(B)$  في بعض الأحيان مصفوفة التباين - التباين، أو مصفوفة التباين والتباين المشترك . وبشكل أكثر تفصيلاً يمكن كتابة  $\text{Cov}(B)$  كالتالي:

$$\text{Cov}(B) = E \left\{ \begin{bmatrix} B_0 - \beta_0 \\ B_1 - \beta_1 \\ \vdots \\ B_k - \beta_k \end{bmatrix} \begin{bmatrix} B_0 - \beta_0 & B_1 - \beta_1 & \cdots & B_k - \beta_k \end{bmatrix} \right\}$$

$$= \begin{bmatrix} E(B_0 - \beta_0)^2 & E(B_0 - \beta_0)(B_1 - \beta_1) & \cdots & E(B_0 - \beta_0)(B_k - \beta_k) \\ E(B_1 - \beta_1)(B_0 - \beta_0) & E(B_1 - \beta_1)^2 & & E(B_1 - \beta_1)(B_k - \beta_k) \\ \vdots & \vdots & \ddots & \vdots \\ E(B_k - \beta_k)(B_0 - \beta_0) & E(B_k - \beta_k)(B_1 - \beta_1) & \cdots & E(B_k - \beta_k)^2 \end{bmatrix}$$

واضح أن المصفوفة  $\text{Cov}(B)$  متماثلة ومن الدرجة  $(k+1)(k+1)$  كما أن العنصر  $i$  على القطر الرئيسي يمثل تباين  $B_i$  والعنصر  $ii'$  فوق القطر الرئيسي يمثل التغاير بين  $B_i, B_{i'}$  حيث  $i \neq i'$ . أي أن:

$$E[B_i - \beta_i]^2 = \text{Var}(B_i) ,$$

$$E[B_i - \beta_i][B_{i'} - \beta_{i'}] = \text{Cov}(B_i, B_{i'}) ,$$

$$i, i' = 0, 1, 2, \dots, k+1.$$

### نظرية (١-٣) (نظرية جلوس-ماركوف)

مقدر المربعات الصغرى  $B$  يمثل أفضل مقدر خطي غير متحيز وله أقل تباين حيث:

$$\text{Cov}(B) = \sigma^2 (X'X)^{-1} . \quad (٢٢-٣)$$

البرهان

اثبتنا من قبل أن المقدّر  $B$  غير متحيز للمعلمة  $\beta$ .

(١) لإثبات (٢٢-٣) ويفرض  $A = (X'X)^{-1}X'$  ومن (١٧-٣) فإن  $B = AY$  ومن (٢١-٣) فإن:

$$\begin{aligned}\text{Cov}(B) &= A\text{Cov}(Y)A' \\ &= \sigma^2 AIA' = \sigma^2 AA' \\ &= \sigma^2 (X'X)^{-1} X'X (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1},\end{aligned}$$

والذي يكمل البرهان.

وعلى ذلك بوضع  $(X'X)^{-1} = C$  حيث:

$$C = \begin{bmatrix} c_{00} & c_{01} & c_{02} & \cdots & c_{0p} \\ c_{10} & c_{11} & c_{12} & \cdots & c_{1p} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ c_{p0} & c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix}$$

و  $p = k+1$  فإن المصفوفة  $\text{Cov}(B)$  يمكن كتابتها على الشكل التالي:

$$\text{Cov}(B) = \sigma^2 C.$$

وعلى ذلك تبين  $B_i$  هو  $\sigma^2 c_{ii}$  والتغاير بين  $B_i, B_{i'}$  هو  $\sigma^2 c_{ii'}$  حيث  $i \neq i'$ .

(ب) ولإثبات أن المتجه  $B$  له أقل تباين نتبع الآتي:

نفرض أن  $B^*$  أي مقدر خطي آخر حيث:

$$B^* = [(X'X)^{-1}X' + D]Y,$$

حيث المصفوفة  $D$  من الدرجة  $p \times n$  وعناصرها ثوابت. وبما أن  $Y = X\beta + \varepsilon$  فإن:

$$B^* = [(X'X)^{-1}X' + D][X\beta + \varepsilon]$$

$$= (X'X)^{-1}X'X\beta + DX\beta + (X'X)^{-1}X'\varepsilon + D\varepsilon \quad (٧٣-٣)$$

وحتى يكون المتجه  $B^*$  مقدر غير منحيز للمتجه  $B$  يجب أن تكون المصفوفة  $DX$  مصفوفة صفرية أي أن:

$$DX = 0$$

يمكن إعادة كتابة المعادلة (٧٣-٣) كالآتي:

$$B^* - \beta = (X'X)^{-1}X'\varepsilon + D\varepsilon$$

وعلى ذلك:

$$\begin{aligned}
 \text{Cov}(B^*) &= E\left\{ [B^* - \beta][B^* - \beta]' \right\} \\
 &= E\left[ (X'X)^{-1}X'\varepsilon + D\varepsilon \right] \left[ (X'X)^{-1}X'\varepsilon + D\varepsilon \right]' \\
 &= E\left[ (X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1} + D\varepsilon\varepsilon'X(X'X)^{-1} + (X'X)^{-1}X'\varepsilon\varepsilon'D' + D\varepsilon\varepsilon'D' \right] \\
 &= \sigma^2(X'X)^{-1} + \sigma^2DX(X'X)^{-1} + \sigma^2(X'X)^{-1}X'D' + \sigma^2DD'. \quad (٢٤-٣)
 \end{aligned}$$

وباستخدام الفرض  $DX = 0$  أو  $X'D' = 0$  فإن المعادلة (٢٤-٣) تصبح :

$$\text{Cov}(B^*) = \sigma^2(X'X)^{-1} + \sigma^2DD'. \quad (٢٥-٣)$$

من (٢٥-٣) يتضح أن المصفوفة  $\text{Cov}(B^*)$  تساوي المصفوفة  $\text{Cov}(B)$  مضاف عليها القيمة  $(\sigma^2DD')$  . وعلى ذلك فإن تباين كل عنصر من عناصر المتجه  $B^*$  أكبر من أو يساوي تباين العنصر المقابل في المتجه  $B$  . وبالتالي لا توجد مقدرات تباينها أقل من تباين مقدرات المربعات الصغرى . أيضا تحت فرض أن  $\varepsilon$  تتبع توزيعا طبيعيا فإن  $B$  أيضا سوف يكون هو أيضا مقدر الإمكان الأعظم .

(ج) لإثبات صفة الخطية للمقدر  $B$  نتبع الآتي :

بما أن مقدر المربعات الصغرى للمعلمة  $\beta$  هو  $B$  حيث :

$$B = (X'X)^{-1}X'Y$$

وبما أن  $(X'X)^{-1}X'$  مصفوفة أرقام ثابتة ، فإن  $B$  دالة خطية لـ  $Y$  ، ومن ثم فإن مقدر المربعات الصغرى  $B$  مقدر خطي .

### (٧-٣) خواص البواقي

١- استقلال البواقي عن المصفوفة  $X$ ، أي أن  $X'e = 0$

البرهان

بإستخدام المصفوفة  $M$  و (١٣-٣) فإننا يمكن التعبير عن  $e$  بدلالة  $Y$  أو  $\varepsilon$  كالآتي :

$$e = Y - HY = MY$$

$$= MX\beta + M\varepsilon = M\varepsilon. \quad (٢٦-٣)$$

وبما أن:  $MX = 0$  من (٣-١٤)

$$X'e = X'M\varepsilon = 0 \quad \text{وعلى ذلك:}$$

أي المتجه الذي كله أصفار. وعلى ذلك إذا كانت المعلمة  $\beta_0$  موجودة في نموذج

الإنحدار المتعدد فإن العمود الأول من المصفوفة  $X$  هو  $1 = (1, \dots, 1)'$

وعلى ذلك:

$$\sum e_i = 1'e = 0$$

(هنا ننظر إلى  $e$  كمصفوفة عشوائية بتكرار المعادلة).

٢- إستقلال البواقي عن القيم المقدرة للمتغير التابع. أي أن:

$$\hat{y}'e = 0$$

**البرهان**

$$\hat{y}'e = b'X'e = 0.$$

٣- القيم المتوقعة لأي عنصر من عناصر متجه البواقي  $e$  تساوي صفراً ، أي أن:

$$E(e) = 0.$$

**البرهان**

$$E(e) = E(M\varepsilon) = ME(\varepsilon) = 0.$$

٤- تباین متجه البواقي يساوي  $\sigma^2 M$  ، أي أن:

$$\text{Cov}(e) = \sigma^2 M$$

**البرهان**

$$\text{Cov}(e) = E[e - E(e)][e - E(e)]'$$

وبما أن  $E(e) = 0$  فإن:

$$\text{Cov}(e) = E(ee')$$

وبما أن  $c = M\varepsilon$  فإن:

$$\text{Cov}(e) = E(M\varepsilon\varepsilon'M')$$

$$= M\sigma^2 M'$$

$$= \sigma^2 MM'$$

$$= \sigma^2 M .$$

وعلى ذلك فإن:

$$\text{Var}(e_j) = \sigma^2 m_{jj} = \sigma^2 [1 - h_{jj}]$$

حيث  $m_{jj}$  ,  $h_{jj}$  العنصر  $jj$  للمصفوفة  $H, M$  على التوالي حيث  $0 \leq h_{jj} \leq 1$  .

التغاير بين  $e_j, e_{j'}$  ,  $j \neq j'$  هو :

$$\text{Cov}(e_j, e_{j'}) = -\sigma^2 h_{jj} .$$

مثال (٢-٣)

يتأثر محصول الفراولة بكمية الأمطار  $x_1$  وكمية السماد المستخدم  $x_2$  .  
يستخدم البيانات في جدول (٢-٣) لتوفيق معادلة إحدار خطي متعدد باستخدام كمية  
الأمطار وكمية السماد كمغيرات مستقلة.

جدول (٢-٣)

$x_1$	$x_2$	$y$
16	510	1000
22	450	450
23	500	1200
13	425	700
17	450	800
25	475	1100
18	515	1050
20	500	1150
21	490	1000
19	510	950
22	525	1300

الحل

لإيجاد معادلة الإحدار المقدرة

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 .$$

وباستخدام البيانات في جدول (٢-٣) فإن المصفوفة  $X$  والمتجه  $y$  يكونان  
على الشكل التالي:

$$X = \begin{bmatrix} 1 & 16 & 510 \\ 1 & 22 & 450 \\ 1 & 23 & 500 \\ 1 & 13 & 425 \\ 1 & 17 & 450 \\ 1 & 25 & 475 \\ 1 & 18 & 515 \\ 1 & 20 & 500 \\ 1 & 21 & 490 \\ 1 & 19 & 510 \\ 1 & 22 & 525 \end{bmatrix} \quad , \quad y = \begin{bmatrix} 1000 \\ 450 \\ 1200 \\ 700 \\ 800 \\ 1100 \\ 1050 \\ 1150 \\ 1000 \\ 950 \\ 1300 \end{bmatrix}$$

المصفوفة  $X'X$  ستكون على الشكل التالي:

$$X'X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 16 & 22 & \dots & 22 \\ 510 & 450 & \dots & 525 \end{bmatrix} \begin{bmatrix} 1 & 16 & 510 \\ 1 & 22 & 450 \\ \vdots & \vdots & \vdots \\ 1 & 22 & 525 \end{bmatrix} ,$$

$$X'y = \begin{bmatrix} 10700 \\ 213250 \\ 5.26525 \times 10^6 \end{bmatrix} .$$

تقديرات المربعات الصغرى سوف نحصل عليها كالتالي:

$$b = (X'X)^{-1} X'y .$$

أي:

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 11 & 216 & 5350 \\ 216 & 4362 & 105410 \\ 5350 & 105410 & 2.6124 \times 10^6 \end{bmatrix}^{-1} \begin{bmatrix} 10700 \\ 213250 \\ 5.26525 \times 10^6 \end{bmatrix}$$

$$= \begin{bmatrix} 23.0157 & -0.0271387 & -0.0460394 \\ -0.0271387 & 0.00922992 & -0.000316848 \\ -0.0460394 & -0.000316848 & 0.000107453 \end{bmatrix} \begin{bmatrix} 10700 \\ 213250 \\ 5026525 \times 10^6 \end{bmatrix}$$

$$= \begin{bmatrix} -1928.24 \\ 9.61221 \\ 5.57653 \end{bmatrix}.$$

وعلى ذلك معادلة الإنحدار المتعدد المقدرة سوف تكون على الشكل:

$$\hat{y} = -1928.24 + 9.61221x_1 + 5.57653x_2.$$

ويمكن تفسير التقدير  $b_0$  على أنه يمثل القيمة المقدرة للمتغير التابع عندما تكون قيم المتغيرات المستقلة تساوي الصفر. وفي الواقع فإن هذا التفسير غير صحيح في كل الحالات. فبإتباع هذا التفسير نجد أن محصول القراولة يكون سالبا  $-1928.24$  عندما تكون كمية الأمطار تساوي صفر وكمية السماد يساوي صفر وهذا غير منطقي. كما أن مشاهدات العينة لا تحتوي على قيم صفرية لكل من كمية الأمطار وكمية المحصول. يعطي جدول (٣-٣) المشاهدات ( $y$ ) القيم المقدرة ( $\hat{y}$ ) والبقايا  $e$ .

جدول (٣-٣)

$y$	$\hat{y}$	$e = y - \hat{y}$
1000	1069.58	-69.5826
450	792.664	-342.664
1200	1081.1	118.897
700	566.741	133.259
800	744.603	55.3967
1100	960.914	139.086
1050	1116.69	-66.6896
1150	1052.27	97.7338
1000	1006.11	-6.1131
950	1098.42	-148.419
1300	1210.9	89.0963



(٨-٣) صيغة أخرى للحصول على تقديرات المربعات  
الصغرى لمعامل نموذج الانحدار الخطي  
المتعدد

وكما في الانحدار البسيط فإن هناك صيغة مكافئة للحصول على تقديرات المربعات الصغرى لمعامل النموذج الخطي المتعدد تعتمد على مجموع المربعات المصحح وحواصل الضرب. للوصول إلى ذلك ، ليكن  $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k)'$  متجه من الدرجة  $k \times 1$  من متوسطات العينة للمتغيرات المستقلة.

وإذا كانت  $X_0$  ترمز لمصفوفة من الدرجة  $n \times k$  من البيانات الأصلية مطروح منها المتوسطات وعلى ذلك العنصر  $(i, j)$  من  $X_0$  هو  $x_{ij} - \bar{x}_j$ . وبنفس الشكل يمكن تعريف  $y_0$  متجه من الدرجة  $n \times 1$  حيث العنصر رقم  $j$  هو  $y_j - \bar{y}$ . وعلى ذلك المصفوفة  $X_0'X_0$  والمصفوفة  $X_0'y_0$  تأخذان الشكل التالي:

$$X_0'X_0 = \begin{bmatrix} \sum (x_{1j} - \bar{x}_1)^2 & \dots & \sum (x_{1j} - \bar{x}_1)(x_{kj} - \bar{x}_k) \\ \vdots & & \vdots \\ \sum (x_{1j} - \bar{x}_1)(x_{kj} - \bar{x}_k) & \dots & \sum (x_{kj} - \bar{x}_k)^2 \end{bmatrix}$$

$$X_0'y_0 = \begin{bmatrix} \sum (x_{1j} - \bar{x}_1)(y_j - \bar{y}) \\ \vdots \\ \sum (x_{kj} - \bar{x}_k)(y_j - \bar{y}) \end{bmatrix}$$

إن فالمعادلات الطبيعية باستخدام مجموع المربعات ومجموع حاصل الضرب المصحح هما:

$$X_0'X_0b^* = X_0'y_0$$

وباستخدام طريقة المربعات الصغرى فإن تقدير  $B^*$  هو:

$$b^* = (X_0'X_0)^{-1}X_0'y_0 = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}$$

أما  $b_0$  فيقدر كالاتي:

$$b_0 = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - \dots - b_k\bar{x}_k$$

حيث:

$$b = \begin{pmatrix} b_0 \\ b^* \end{pmatrix}$$

وعند مقارنة الطريقتين نجد أن:

١. درجة المصفوفة  $X'X$  هو  $(k+1) \times (k+1)$  بينما المصفوفة  $X_0'X_0$  من الدرجة  $(k \times k)$ .

٢. عناصر المصفوفة  $(X_0'X_0)^{-1}$  هي نفسها عناصر المصفوفة  $(X'X)^{-1}$  بعد حذف الصف الأول والعمود الأول من  $(X'X)^{-1}$ .

٣. عناصر المتجه  $b$  هي نفسها قيم عناصر  $b^*$  بعد حذف  $b_0$  من  $b$ .

(٩-٣) تقدير  $\sigma^2$

كما في نموذج الانحدار الخطي البسيط فإن تقدير التباين  $\sigma^2$  يمكن الحصول عليه من مجموع مربعات البواقي:

$$\begin{aligned} SSE &= \sum_{j=1}^n (y_j - \hat{y}_j)^2 \\ &= \sum_{j=1}^n e_j^2 \\ &= e'e. \end{aligned}$$

وبوضع  $e = (y - Xb)$  في SSE نحصل على:

$$\begin{aligned} SSE &= (y - Xb)' (y - Xb) \\ &= y'y - b'X'y - y'Xb + b'X'Xb. \end{aligned} \quad (٢٧-٣)$$

وبما أن:

$$X'Xb = X'y$$

فإن (٢٧-٣) تصبح:

$$SSE = y'y - b'X'y. \quad (٢٨-٣)$$

درجات الحرية لمجموع مربعات البواقي تساوي  $n - p$  وذلك لوجود  $p = (k + 1)$  معلمة تقدر في نموذج الانحدار. متوسط مجموع المربعات هو (التقدير للتباين  $\sigma^2$ ):

$$MSE = \frac{SSE}{n - p} = s^2. \quad (٢٩-٣)$$

وكما في نموذج الانحدار الخطي البسيط فإن  $s^2$  تعتمد على النموذج.

#### نظرية (٢-٣)

تحت فروض جاوس - ماركوف ، فإن  $S^2$  هو مقدر غير متحيز للمعلمة  $\sigma^2$ .

#### البرهان

$$e'e = y'y - b'X'y = \sum e_i^2$$

حيث  $\sum e_i^2$  يساوي مجموع العناصر القطرية (أثر المصفوفة trace والذي يرمز له بالرمز  $\text{tr}$ ) للمصفوفة  $ee'$  أي أن :

$$e'e = \text{tr}(ee')$$

إذن:

$$\begin{aligned} E(e'e) &= E[\text{tr}(ee')] \\ &= \text{tr} E[(e'e)] \\ &= \text{tr} E(M\epsilon\epsilon'M) \\ &= \text{tr}(\sigma^2 M) \\ &= \sigma^2 \text{tr}(I - X(X'X)^{-1}X') \\ &= \sigma^2 [\text{tr} I - \text{tr}(X'X)^{-1}X'X] \\ &= \sigma^2 [\text{tr} I_{n \times n} - \text{tr} I_{p \times p}] \\ &= \sigma^2 [n - p]. \end{aligned}$$

وذلك لأن مصفوفة الوحدة  $I$  يتكون عناصر قطرها من الواحد الصحيح وعلى ذلك:

$$E(S^2) = E\left(\frac{e'e}{n - p}\right) = \frac{\sigma^2 [n - p]}{[n - p]} = \sigma^2.$$

مثال (٣-٣)

في دراسة عن العلاقة بين إمتصاص الماء في دقيق القمح والخواص المختلفة للدقيق ونحت فرض نموذج إنداز خطي متعدد ثم الحصول على البيانات في جدول (٣-٤) حيث  $Y$  تمثل كمية إمتصاص الماء و  $x_1$  كمية البروتين و  $x_2$  كمية النشا الذي يتعرض للفقد (التحطم مفاًس بوحدات Farrand) والمطلوب إيجاد معادلة الإنداز المقدرة وتقدير تباين الخطأ  $\sigma^2$ .

جدول (٣-٤)

$x_1$	$x_2$	$y$
8.5	2	30.9
8.9	3	32.7
10.6	3	36.7
10.2	20	41.9
9.8	22	40.9
10.8	20	42.9
11.6	31	46.3
12	32	47.2
12.5	31	44
10.4	28	47.7
1.2	36	43.9
11.9	28	46.8
11.3	30	46.2
13	27	47
12.9	24	46.8
12	25	45.9
12.9	28	48.8
13.1	28	46.2
11.4	32	47.8
13.2	28	49.2

الحل

المصفوفات  $X$  و  $y$  هما:

$$X = \begin{bmatrix} 1 & 8.5 & 2 \\ 1 & 8.9 & 3 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 13.2 & 28 \end{bmatrix} \quad y = \begin{bmatrix} 30.9 \\ 32.7 \\ \vdots \\ \vdots \\ 49.2 \end{bmatrix},$$

المصفوفة  $X'X$  هي:

$$X'X = \begin{bmatrix} 1 & 1 & \dots & \dots & 1 \\ 8.5 & 8.9 & \dots & \dots & 13.2 \\ 2 & 3 & \dots & \dots & 28 \end{bmatrix} \cdot \begin{bmatrix} 1 & 8.5 & 2 \\ 1 & 8.9 & 3 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 13.2 & 28 \end{bmatrix}$$

$$= \begin{bmatrix} 20 & 218.2 & 478 \\ 218.2 & 2515.88 & 5271.8 \\ 478 & 5271.8 & 13322 \end{bmatrix}.$$

والمنهج  $X'y$  هو:

$$X'y = \begin{bmatrix} 1 & 1 & \dots & \dots & 1 \\ 8.5 & 8.9 & \dots & \dots & 13.2 \\ 2 & 3 & \dots & \dots & 28 \end{bmatrix} \cdot \begin{bmatrix} 30.9 \\ 32.7 \\ \vdots \\ \vdots \\ 49.2 \end{bmatrix}$$

$$= \begin{bmatrix} 879.8 \\ 9710.06 \\ 21894.8 \end{bmatrix}.$$

قيم  $b$  تعطى من العلاقة التالية:

$$b = (X'X)^{-1} X'y.$$

أو :

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 20 & 218.2 & 478 \\ 218.2 & 2515.88 & 5271.8 \\ 478 & 5271.8 & 13322 \end{bmatrix}^{-1} \begin{bmatrix} 879.8 \\ 9710.06 \\ 21894.8 \end{bmatrix}$$

$$= \begin{bmatrix} 1.12878 & -0.0762961 & -0.0103092 \\ -0.0762961 & 0.00748409 & -0.000224073 \\ -0.0103092 & -0.000224073 & 0.000533635 \end{bmatrix} \begin{bmatrix} 879.8 \\ 9710.06 \\ 21894.8 \end{bmatrix}$$

$$= \begin{bmatrix} 26.5433 \\ 0.63964 \\ 0.438 \end{bmatrix}$$

معادلة الانحدار المقدرة سوف تكون:

$$\hat{y} = 26.5433 + 0.63964x_1 + 0.438x_2.$$

سوف نقدر تباين الخطأ  $\sigma^2$  كالتالي:

$$y'y = \sum_{j=1}^{20} y_j^2 = 39201.9.$$

وعلى ذلك :

$$b'X'y = [26.5433 \quad 0.63964 \quad 0.438] \cdot \begin{bmatrix} 879.8 \\ 9710.06 \\ 21894.8 \end{bmatrix}$$

$$= 39153.6445.$$

مجموع المربعات للبواقي سيكون :

$$\begin{aligned} SSE &= y'y - b'X'y \\ &= 39201.9 - 39153.6445 \\ &= 48.2555. \end{aligned}$$

وعلى ذلك ، تقدير  $\sigma^2$  هو متوسط مجموع مربعات البواقي ، أي أن :

$$\begin{aligned} s^2 &= \frac{SSE}{n-k-1} = \frac{48.2555}{17} \\ &= 2.8386. \end{aligned}$$

### (١٠-٣) فترات ثقة في الإحذار المتعدد

وكما هو الحال في الإحذار الخطي البسيط فإن فترات الثقة لمعاملات الإحذار أو فترة ثقة لمتوسط الإستجابة أو لإستجابة مفردة عند مستويات خاصة من المتغيرات المستقلة تلعب دوراً مهماً في الإحذار الخطي المتعدد.

### (١٠-٣-١) فترات ثقة لمعاملات الإحذار

للحصول على 100%(1-α) فترات ثقة لمعاملات الإحذار  $\beta_i$  حيث  $i = 0, 1, 2, \dots, k$  لابد أن نفترض أن حدود الخطأ  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  مستقلة وتتبع توزيعات طبيعية بمتوسط صفر وتباين  $\sigma^2$  وعلى ذلك  $Y_1, Y_2, \dots, Y_n$  تمثل متغيرات عشوائية تتبع توزيعات طبيعية بمتوسط  $\beta_0 + \sum_{j=1}^k \beta_j x_{ij}$  وتباين  $\sigma^2$ . وبما أن مقدار المربعات الصغرى B يعتبر تركيبة خطية فهذا يعني أن B تتبع التوزيع الطبيعي بمتوسط  $\beta$  ومصفوفة تغاير  $\sigma^2(X'X)^{-1}$ . وهذا يعني أن التوزيع الهامشي لأي مقدار  $B_i$  يتبع التوزيع الطبيعي بمتوسط  $\beta_i$  وتباين  $\sigma^2 c_{ii}$  حيث  $c_{ii}$  هو العنصر رقم  $i$  ( $i = 0, 1, \dots, k$ ) على القطر الرئيسي للمصفوفة  $(X'X)^{-1}$ . أي أن :

$$B_i \sim N(\beta_i, \sigma^2 c_{ii}), \quad i = 0, 1, 2, \dots, k$$

$$Z = \frac{B_i - \beta_i}{\sqrt{\sigma^2 c_{ii}}}$$

يتبع التوزيع الطبيعي بمتوسط صفر وتباين يساوي الواحد الصحيح، أي أن  $Z \sim N(0, 1)$ .

وبما أن الإحصاء :

$$V = \frac{(n-p)S^2}{\sigma^2}$$

يتبع توزيع مربع كاي بدرجات حرية  $n-p$  حيث  $S^2$  هو مقدار للمعلمة  $\sigma^2$ . وعلى ذلك فإن الإحصاء :

$$T = \frac{Z}{\sqrt{\frac{V}{n-p}}} = \frac{B_i - \beta_i}{\sqrt{S^2 c_{ii}}}$$

يتم توزيع  $t$  بدرجات حرية  $n-p$ .

وعلى ذلك  $100(1-\alpha)\%$  فترة ثقة لمعامل الانحدار  $\beta_i$  حيث  $i = 0, 1, 2, \dots, k$  هو:

$$b_i - t_{\alpha/2}(n-p)\sqrt{s^2 c_{ii}} \leq \beta_i \leq b_i + t_{\alpha/2}(n-p)\sqrt{s^2 c_{ii}}. \quad (٣-٣)$$

حيث  $\sqrt{s^2 c_{ii}}$  هو تقدير للانحراف المعياري للمقدر  $B_i$ .

مثال (٣-٤)

أوجد 95% فترة ثقة للمعلمة  $\beta_1$  في مثال (٣-٢). التقدير بنقطة للمعلمة  $\beta_1$

هو  $b_1 = 9.61221$  والعنصر  $c_{11}$  في المصفوفة  $(X'X)^{-1}$  هو  $c_{11} = 0.00922992$  و  $s^2 = 27571.5$  و  $t_{0.025}(8) = 2.306$  وعلى ذلك من (٣-٣) فإن 95% فترة ثقة للمعلمة  $\beta_1$  هي:

$$b_1 - t_{0.025}(8)\sqrt{s^2 c_{11}} \leq \beta_1 \leq b_1 + t_{0.025}(8)\sqrt{s^2 c_{11}}.$$

أي أن :

$$9.61221 - (2.306)\sqrt{(27571.5)(0.00922992)} \leq$$

$$\beta_1 \leq 9.61221 + (2.306)\sqrt{(27571.5)(0.00922992)}.$$

والتي تختصر إلى:

$$9.61221 - (2.306)(15.9525) \leq \beta_1 \leq 9.61221 + (2.306)(15.9525).$$

وعلى ذلك 95% فترة ثقة للمعلمة  $\beta_1$  هي :

$$-27.1744 \leq \beta_1 \leq 46.3988.$$

٣-١٠-٢) فترة ثقة لمتوسط الإستجابة

الآن إهتمامنا سوف يكون في الحصول على  $100(1-\alpha)\%$  فترة ثقة لمتوسط الإستجابة  $\mu_Y|x_{10}, x_{20}, \dots, x_{k0}$  وذلك لفئة من الظروف التالية:

$$\underline{x}_0' = [1, x_{10}, x_{20}, \dots, x_{k0}]$$

وعلى ذلك القيمة المتنبأ بها للمتغير  $Y_0$  عند النقطة  $\underline{x}_0'$  هي  $\hat{y}_0 = \underline{x}_0' b$  ويمكن إثبات أنه تحت فروض جاوس - ماركوف فإن  $\hat{y}_0$  مقدر غير متحيز

للمعلمة  $\mu_Y|x_{10}, x_{20}, \dots, x_{k0}$



البرهان

$$E(\hat{Y}_0) = E(\underline{x}_0' B) = \underline{x}_0' E(B)$$

$$= \underline{x}_0' \beta = \mu_{Y|x_{10}, x_{20}, \dots, x_{k0}}$$

وعلى ذلك لإيجاد  $100\%(1-\alpha)$  فترة ثقة لمتوسط الإستجابة  $\mu_{Y|x_{10}, x_{20}, \dots, x_{k0}}$  فإننا نحتاج إلى اشتقاق تباين المقدّر  $\hat{Y}_0$  كالتالي:

$$\text{Var}(\hat{Y}_0) = \underline{x}_0' \text{Cov}(B) \underline{x}_0$$

$$= \sigma^2 \left[ \underline{x}_0' (X'X)^{-1} \underline{x}_0 \right].$$

وبما أن  $Y_1, Y_2, \dots, Y_n$  متغيرات عشوائية تتبع توزيعات طبيعية فإن:

$$\hat{Y}_0 \sim (\underline{x}_0' \beta, \sigma^2 [\underline{x}_0' (X'X)^{-1} \underline{x}_0]).$$

وعلى ذلك الإحصاء:

$$Z = \frac{\hat{Y}_0 - \underline{x}_0' \beta}{\sigma \sqrt{\underline{x}_0' (X'X)^{-1} \underline{x}_0}}$$

يتبع التوزيع الطبيعي القياسي بمتوسط صفر وتباين يساوي الواحد الصحيح ، أي أن  $Z \sim N(0,1)$ . أيضاً الإحصاء:

$$V = \frac{(n-p)S^2}{\sigma^2}$$

يتبع توزيع مربع كاي بدرجات حرية  $(n-p)$  حيث  $S^2$  مقدّر للمعلمة  $\sigma^2$ .  
وعلى ذلك الإحصاء :

$$T = \frac{Z}{\sqrt{\frac{V}{n-p}}} = \frac{\hat{Y}_0 - \mu_{Y|x_{10}, x_{20}, \dots, x_{k0}}}{S \sqrt{\underline{x}_0' (X'X)^{-1} \underline{x}_0}}$$

يتبع توزيع  $t$  بدرجات حرية  $n-p$ . وعلى ذلك  $100\%(1-\alpha)$  فترة ثقة للمعلمة  $\mu_{Y|x_{10}, x_{20}, \dots, x_{k0}}$  تكون كالتالي:

$$\hat{y}_0 - t_{\alpha/2}(n-p)\sqrt{s^2 \underline{x}_0' (X'X)^{-1} \underline{x}_0} \leq \mu_Y | x_{10}, x_{20}, \dots, x_{k0}$$

$$\leq \hat{y}_0 + t_{\alpha/2}(n-p)\sqrt{s^2 \underline{x}_0' (X'X)^{-1} \underline{x}_0} \quad (٣١-٣)$$

حيث  $t_{\alpha/2}(n-p)$  قيمة تستخرج من جدول توزيع  $t$  من الملحق (١) بدرجات حرية  $n-p$ .

الكمية  $\sqrt{s^2 \underline{x}_0' (X'X)^{-1} \underline{x}_0}$  تسمى الخطأ المعياري للتنبؤ وعادة تظهر في المخرجات لحزم الحاسب الآلي الجاهزة.

باستخدام مثال (٣-٢) أوجد 95% فترة ثقة لمتوسط الإستجابة عندما  $x_1 = 17, x_2 = 400$ .

الحل

$$\underline{x}_0 = \begin{bmatrix} 1 \\ 17 \\ 400 \end{bmatrix}.$$

وعلى ذلك القيمة المقدرة عند هذه النقطة يمكن الحصول عليها كالتالي:

$$\hat{y}_0 = \underline{x}_0' b = [1 \quad 17 \quad 400] \begin{bmatrix} -1928.24 \\ 9.61221 \\ 5.57653 \end{bmatrix}$$

$$= 465.777.$$

تباين  $\hat{y}_0$  يقدر من الصيغة التالية:

$$s^2 \underline{x}_0' (X'X)^{-1} \underline{x}_0 = 27571.5 [1 \quad 17 \quad 400] \times \begin{bmatrix} 23.0157 & -0.0271387 & -0.0460394 \\ -0.0271387 & 0.00922992 & -0.000316848 \\ -0.0460394 & -0.000316848 & 0.000107453 \end{bmatrix} \begin{bmatrix} 1 \\ 17 \\ 400 \end{bmatrix} = 22394.5.$$

قيمة  $t$  المستخرجة من جدول توزيع  $t$  من الملحق (١) عند درجات حرية 8 هي  $t_{0.025}(8) = 2.306$  من (٣١-٢) فإن 95% فترة ثقة لمتوسط الإستجابة  $\mu_Y | 17, 400$  هي:

$$465.777 - 2.306\sqrt{22394.5} \leq \mu_{Y|7,400} \leq 465.777 + 2.306\sqrt{22394.5}.$$

والتي تختزل إلى:

$$120.688 \leq \mu_{Y|7,400} \leq 810.865.$$

### (٣-١٠-٣) فترة ثقة لمشاهدة مستقبلية

أوجدنا سابقاً في الفصل الأول التنبؤ بفترة ثقة لمشاهدة مستقبلية في حالة الإنحدار البسيط. الآن سوف نتناول حالة الإنحدار المتعدد. ليكن  $y_0$  مشاهدة مستقبلية (مشاهدة جديدة) عند النقطة  $x_0$  من المتغيرات المستقلة، فإنه يمكن الحصول على 100%  $(1-\alpha)$  فترة ثقة لمشاهدة مستقبلية إذا علمنا التوزيع العيني للمتغير  $(\hat{Y}_0 - Y_0)$  حيث  $Y_0 = \mathbf{x}_0' \mathbf{B} + \varepsilon_0$  و  $\hat{Y}_0 = \mathbf{x}_0' \mathbf{B}$  بما أن  $Y_0$  مستقلة عن  $Y_1, Y_2, \dots, Y_n$  وتحت فروض جاوس - ماركوف فإن:

$$\begin{aligned} E(\hat{Y}_0 - Y_0) &= E(\hat{Y}_0) - E(Y_0) \\ &= \mathbf{x}_0' \boldsymbol{\beta} - \mathbf{x}_0' \boldsymbol{\beta} \\ &= 0, \end{aligned}$$

$$\begin{aligned} \text{Var}(\hat{Y}_0 - Y_0) &= \text{Var}(Y_0) + \text{Var}(\hat{Y}_0) \\ &= \sigma^2 + \sigma^2 \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0 \\ &= \sigma^2 [1 + \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0] . \end{aligned}$$

أي أن :

$$\hat{Y}_0 - Y_0 \sim N(0, \sigma^2 [1 + \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0])$$

الإحصاء :

$$Z = \frac{\hat{Y}_0 - Y_0}{\sqrt{\sigma^2 [1 + \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0]}}$$

يتبع التوزيع الطبيعي القياسي بمتوسط صفر وتباين يساوي الواحد الصحيح،  
الإحصاء :

$$V = \frac{(n-p)s^2}{n^2}$$

يتبع توزيع مربع كاي بدرجات حرية  $n-p$  حيث  $S^2$  مقدر للمعلمة  $\sigma^2$  . وعلى ذلك الإحصاء:

$$T = \frac{Z}{\sqrt{\frac{V}{n-p}}} = \frac{\hat{Y}_0 - Y_0}{\sqrt{S^2(1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0)}}$$

يتبع توزيع  $t$  بدرجات حرية  $n-p$  . وعلى ذلك  $100(1-\alpha)\%$  فترة ثقة لإستجابة مفردة  $y_0$  تعطى كالتالي:

$$\begin{aligned} \hat{y}_0 - t_{\alpha/2}(n-p)\sqrt{s^2(1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0)} &\leq y_0 \\ \leq \hat{y}_0 + t_{\alpha/2}(n-p)\sqrt{s^2(1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0)} \end{aligned} \quad (٣٢-٣)$$

حيث  $t_{\alpha/2}(n-p)$  قيمة تستخرج من جدول توزيع  $t$  من الملحق (١) . الفترة السابقة تعتبر تعميم لفترة التنبؤ لملاحظة مستقبلية في حالة نموذج الإنحدار الخطي البسيط.

باستخدام البيانات في المثال (٣-٢) أوجد 95% فترة ثقة عندما  $x_1 = 17, x_2 = 400$  .

**الحل**

من نتائج مثال (٣-٢) فإن 95% فترة ثقة للإستجابة  $y_0$  عندما  $x_1 = 17, x_2 = 400$  يمكن الحصول عليها كالتالي:

$$\begin{aligned} \underline{x}_0' &= [1, 17, 400] \\ \hat{y}_0 &= \underline{x}_0' b = 465.777. \end{aligned}$$

أيضاً من المثال (٣-٢) فإن:

$$\underline{x}_0'(X'X)^{-1}\underline{x}_0 = 0.812233.$$

وعلى ذلك من (٣-٣) فإن فترة ثقة لإستجابة مفردة  $y_0$  تعطى كالتالي:

$$465.777 - 2.306\sqrt{49966} \leq y_0 \leq 465.777 + 2.306\sqrt{49966}$$

والتي تختزل إلى:

$$-49.6857 \leq y_0 \leq 981.25.$$

(٣-١-٤) فترة ثقة لدالة خطية لعدة معاملات إتحدار

بفرض أن نموذج الإتحدار الخطي المتعدد هو:

$$Y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \epsilon_j$$

وبفرض أن  $\psi = TB$  دالة لعدة معاملات إتحدار حيث  $T$  هي:

$$T = [0, -1, 1]$$

أي أن الدالة الخطية المقدرة (ولتكن  $\hat{\psi}$ ) هي:

$$\hat{\psi} = b_2 - b_1$$

للحصول على  $100\%(1-\alpha)$  فترة ثقة لـ  $\psi$  نتبع الآتي:

$$\hat{\psi} - t_{\alpha/2}(n-p)S_{\hat{\psi}} \leq \psi \leq \hat{\psi} + t_{\alpha/2}(n-p)S_{\hat{\psi}} \quad (٣-٣)$$

حيث:

$$S_{\hat{\psi}} = \sqrt{s^2[c_{11} + c_{22} - 2c_{12}]} \quad (٣-٤)$$

بالرجوع للمثال (٣-٢) أوجد 95% فترة ثقة لـ  $\psi = \beta_2 - \beta_1$

الحل

$$\hat{\psi} = b_2 - b_1 = 5.57653 - 9.61221 = -4.03568$$

حيث:

$$b_2 = 5.57653 \quad , \quad b_1 = 9.61221$$

ومن (٣-٤) فإن :

$$S_{\hat{\psi}} = \sqrt{27571.5[0.00922992 + .000107 - 2(-0.0003168)]}$$

$$= 16.58,$$

$$t_{\alpha/2}(n-p) = t_{0.025}(8) = 2.306$$

وعلى ذلك 95% فترة ثقة للمعلمة  $\psi$  من (٣-٣) هي:

$$\hat{\psi} - t_{\alpha/2}(8)16.58 \leq \beta_2 - \beta_1 \leq \hat{\psi} + t_{\alpha/2}(8)16.58$$

بالتعويض فإننا نحصل على:

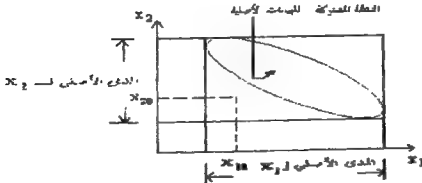
$$-42.269 \leq \beta_2 - \beta_1 \leq 34.19787.$$

### (١١-٣) تقديرات أو تنبؤات خارج مجال النموذج

ينبغي التحذير من القيام بالتنبؤ بإستجابة جديدة أو بتقدير متوسط الإستجابة عند نقطة معطاة  $x_{10}, x_{20}, \dots, x_{k0}$  وذلك خارج المنطقة التي تحتوي على المشاهدات الأصلية. في نموذج الإتحاد المتحد من السهل تحديد المنطقة التي تحتوي على المشاهدات الأصلية حيث مستويات المتغيرات المستقلة  $(x_{1j}, x_{2j}, \dots, x_{kj})$  و  $j=1, 2, \dots, n$  تمثل المنطقة التي تحتوي على البيانات. فعلى سبيل المثال في شكل (٣-٤) والذي يوضح المنطقة التي تحتوي على البيانات الأصلية وذلك لمتغيرين مستقلين. يتضح من شكل (٣-٤) أن النقطة  $(x_{10}, x_{20})$  تقع في مجال كل من المتغيرين  $x_1, x_2$ . ولكن خارج المنطقة الممثلة للبيانات الأصلية. وعلى ذلك فإن التنبؤ لمشاهدة جديدة أو تقدير متوسط الإستجابة عند هذه النقطة سوف يكون خارج مجال النموذج. سوف نستخدم طريقة لتحديد المنطقة التي يغطيها  $x_1, x_2$  معا والتي تعتمد على مصفوفة القبة  $H$ . حيث  $h_{jj}$  تمثل عنصر على القطر الرئيسي للمصفوفة  $H$ . الفئة من نقاط  $x$  ( ليس من الضروري نقاط البيانات المستخدمة في توفيق النموذج ) لابد أن تحقق الشرط التالي :

$$x'(X'X)^{-1}x \leq h_{\max}.$$

عند الإهتمام بالتنبؤ أو بالتقدير عند النقطة  $x_0 = [1, x_{10}, x_{20}, \dots, x_{k0}]$  فإننا نحسب القيمة  $h_{00} = x_0'(X'X)^{-1}x_0$ . النقاط التي تحقق الشرط  $h_{00} > h_{\max}$  تمثل نقاط خارج من النموذج.



شكل (٣-٤)

### (١٢-٣) إختبارات الفروض

في مشاكل الإتحدار المتعدد يوجد عديد من إختبارات الفروض التي تخص معالم النموذج والمقيدة في إختيار جودة النموذج. في هذا البند سوف نتناول بعض إختبارات الفروض المهمة. سوف نضيف إلى فروض جاوس - ماركوف فرض الإعتدال والإستقلال لحدود الخطأ  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  والتي نتاولناها في البند (٣-٥).

### (١-١٢-٣) إختبار يخص جميع معاملات الإتحدار الجزئية

يقدر إختبار المعنوية فيما إذا كان هناك علاقة خطية بين متغير الإستجابة  $Y$  وأي من المتغيرات المستقلة  $x_1, x_2, \dots, x_k$  وبعبارة أخرى هل هناك تأثير معنوي لجميع (أو بعض) المتغيرات المستقلة على المتغير  $Y$ .  
الفرض المناسب هو :

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

ضد الفرض البديل ليست كل  $\beta_i (i=1, \dots, k)$  تساوي صفر.  $H_1$ : رفض  $H_0$ . يعني أنه على الأقل يوجد واحد من  $x_1, x_2, \dots, x_k$  يرتبط معنويًا بالنموذج. يعتبر هذا الإختبار تعميم للإختبار المستخدم في الإتحدار الخطي البسيط. مجموع المربعات الكلي يجزئ إلى مجموع المربعات التي تعود إلى الإتحدار ومجموع المربعات التي تعود إلى الخطأ (مجموع مربعات البواقي). أي أن :

$$SYY = SSR + SSE$$

وعندما  $H_0$  صحيح فإن  $\frac{SSR}{\sigma^2} \sim \chi^2_{(k)}$  حيث درجات الحرية لـ  $\chi^2$  تساوي عدد

المتغيرات المستقلة في النموذج. أيضاً يمكن إثبات أن  $\frac{SSE}{\sigma^2} \sim \chi^2_{(n-k-1)}$  وأن  $SSR$  و  $SSE$  مستقلين. لإختبار فرض العدم  $H_0$  فإننا نحسب قيمة للإحصاء :

$$F = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE} \quad (٣٥-٣)$$

نرفض  $H_0$  عندما تكون قيمة  $F$  المحسوبة أكبر من القيمة الجدولية  $F_{\alpha}[k, n-k-1]$  والمستخرجة من جدول  $F$  في الملحق (٢) عند  $\alpha = 0.05$  أو الملحق (٣) عند  $\alpha = 0.01$ . نحسب مجاميع المربعات كالتالي :

مجموع المربعات للخطأ سيكون :

$$SSE = y'y - b'X'y .$$

مجموع المربعات الكلي سيكون :

$$SYY = \sum y_j^2 - \frac{(\sum_{j=1}^n y_j)^2}{n} = y'y - \frac{(\sum_{j=1}^n y_j)^2}{n} .$$

يمكن كتابة SSE كالتالي :

$$SSE = \left[ y'y - \frac{(\sum_{j=1}^n y_j)^2}{n} \right] - \left[ b'X'y - \frac{(\sum_{j=1}^n y_j)^2}{n} \right]$$

$$= SYY - SSR$$

وعلى ذلك مجموع المربعات للإنحدار سيكون :

$$SSR = b'X'y - \frac{(\sum_{j=1}^n y_j)^2}{n} .$$

النتائج السابقة يمكن وضعها في جدول تحليل التباين المعطى في جدول (٥-٣)

جدول (٥-٣)

S.O.V	df	SS	MS	F
الإنحدار	k	SSR	MSR=SSR/k	MSR/MSE
الخطأ	n-k-1	SSE	MSE=SSE/n-k-1	
الكلي	n-1	SYY		

للمثال (٢-٣) إختبر معنوية الانحدار ؟



$$\begin{aligned} SYY &= y'y - \frac{(\sum y_j)^2}{n} \\ &= 11000000 - \frac{(10700)^2}{11} \\ &= 591818 , \end{aligned}$$

$$\begin{aligned} SSR &= b'X'y - \frac{(\sum y_j)^2}{n} \\ &= 10779427.8 - 10408181.8 \\ &= 371246 , \end{aligned}$$

$$\begin{aligned} SSE &= y'y - b'X'y \\ &= SYY - SSR \\ &= 591818 - 371246 \\ &= 220572 . \end{aligned}$$

جدول تحليل التباين معطى في جدول (٦-٣)

جدول (٦-٣)

S.O.V	df	SS	MS	F
الإنحدار	2	371246	185623	6.73243
الخطأ	8	220572	27571.5	-
الكلي	10	591818	-	-

لإختبار فرض العدم :

$$H_0 : \beta_1 = \beta_2 = 0 .$$

سوف نحسب قيمة للإحصاء F كالتالى:

$$F = \frac{MSR}{MSE} = \frac{185623}{27571.5} = 6.73243$$

وبما أن قيمة  $F$  المحسوبة تزيد عن قيمة  $F$  الجدولية المستخرجة من جدول توزيع  $F$  في الملحق (٢) عند  $\alpha = 0.05$  و  $F_{0.05}[2,8] = 4.46$  ، فإننا نستنتج أن  $Y$  ترتبط مع  $x_1$  و  $x_2$  . في بعض الأحيان فإن هذا لايعطي أن هذه العلاقة منسبة للتنبؤ بالمتغير التابع كدالة في  $x_2, x_1$  .

مجاميع المربعات في تحليل التباين بدلالة المصفوفات هي:

$$SYY = y'y - \left(\frac{1}{n}\right)y'y = y' \left[ I - \left(\frac{1}{n}\right)J \right] y,$$

$$SSE = y'y - b'X'y = y'(I - H)y,$$

$$SSR = b'X'y - \left(\frac{1}{n}\right)y'y$$

$$= y' \left[ H - \left(\frac{1}{n}\right)J \right] y.$$

حيث  $J$  مصفوفة جميع عناصرها الواحد الصحيح ومن الرتبة  $n \times n$  .  
ويمكن التعبير عن جدول تحليل التباين بصورة أخرى والمعطى في جدول (٧-٣).

جدول (٧-٣)

S.O.V	df	SS
إنحدار $\beta_0$	1	$\frac{(\sum_{j=1}^n y_j)^2}{n}$
إنحدار $\beta_1, \beta_2   \beta_0$	(k)	$b'X'y - \frac{(\sum_{j=1}^n y_j)^2}{n}$
الخطأ	(n-k-1)	(بالطرح)
المجموع الكلي الغير مصحح	(n)	$\sum_{j=1}^n y_j^2$

يمكن كتابة جدول تحليل التباين بشكل آخر كما هو معطى في جدول (٨-٣) وذلك في حالة  $k$  من المتغيرات المستقلة.

جدول (٨-٣)

S.O.V	df	SS	MS
إنحدار $\beta_0, \beta_1, \dots, \beta_k$	$(k+1)$	$SSR(\beta_1, \dots, \beta_k, \beta_0)$ $= b'X'y$	$b'X'y / (k+1)$
الخطأ	$(n-k-1)$	$y'y - b'X'y$	$(y'y - b'X'y) / (n-k-1)$
المجموع	$n$	$y'y$	

(٣-١٢-٢) معامل التحديد المتعدد

### Coefficient of Multiple Determination

يقيس معامل التحديد نسبة التباين أو التغير في المتغير التابع  $Y$  التي تفسرها المتغيرات المستقلة  $x_1, x_2, \dots, x_k$ ، أي أنه يقيس نسبة التباين في  $Y$  التي يمكن تفسيرها بمعادلة الإنحدار المتعدد المقدرة. وكما في حالة الإنحدار البسيط فإن معامل التحديد في حالة  $k$  من المتغيرات المستقلة يحسب من الصيغة التالية:

$$R^2 = \frac{SSR}{SY Y} = 1 - \frac{SSE}{SY Y}$$

حيث يستخدم معامل التحديد المتعدد في تقييم جودة توفيق خط إنحدار العينة لتقييم مشاهدات المتغير التابع  $Y$ . كما هو الحال في الإنحدار الخطي البسيط فإن  $0 \leq R^2 \leq 1$ . في بعض الأحيان فإن كبر  $R^2$  لا يعني بالضرورة أن نموذج الإنحدار جيد. إن إضافة متغير مستقل إلى النموذج دائما يؤدي إلى زيادة  $R^2$  بصرف النظر عن ما إذا كان هذا المتغير ضروري للنموذج أم لا. وعلى ذلك من الممكن للنماذج التي بها قيم  $R^2$  عالية أن تكون نماذج رديئة. الجذر التربيعي لـ  $R^2$  هو معامل الارتباط المتعدد بين  $Y$  وفترة المتغيرات المستقلة  $x_1, x_2, \dots, x_k$ . أي لن:

$$R = +\sqrt{R^2}$$

أي أن معامل الارتباط المتعدد يأخذ قيمة غير سالبة وهو يختلف في هذه الصفة عن معامل الارتباط البسيط الذي يمكن أن يأخذ قيمة سالبة ، أي أن :

$$0 \leq R \leq 1$$

حيث  $R$  مقياس لقوة العلاقة الخطية بين  $Y$  و  $x_1, x_2, \dots, x_k$  .

وبأخذ  $R^2$  القيمة 0 عندما يكون جميع المقادير  $b_i = 0$  ،  $i = 1, 2, \dots, p-1$  مساوية للصفر وبأخذ  $R^2$  القيمة 1 عندما تقع جميع المشاهدات  $y$  على سطح الإستجابة التوفيقي مباشرة، أي عندما يكون  $\hat{y}_i = y_i$  لجميع قيم  $i$  . ولأن معامل التحديد  $R^2$  دالة تزايدية لعدد المتغيرات المستقلة فإضافة أي متغير مستقل لنموذج الإنحدار تزيد قيمة المعامل بغض النظر عن مساهمة هذا المتغير في تفسير تباين المتغير التابع، ولذا ولغرض الحصول على مقياس أفضل لقياس مدى قابلية مجاميع مختلفة من المتغيرات لتحليل العلاقة قيد الدراسة وفي نفس الوقت إذ يأخذ في الاعتبار عدد المتغيرات في النموذج فإنه يتم حساب ما يعرف بمعامل التحديد المعدل والذي يأخذ الصيغة التالية:

$$\bar{R}^2 = 1 - \frac{SSE/(n-k-1)}{SYY/(n-1)}$$

حيث  $k$  عدد المتغيرات المستقلة، ويمكن بسهولة اشتقاق صيغة للعلاقة بين  $R^2$ ،  $\bar{R}^2$  كالتالي:

$$\begin{aligned} \bar{R}^2 &= 1 - \frac{SSE/(n-k-1)}{SYY/(n-1)} \\ &= 1 - \frac{(SYY - SSR)/(n-k-1)}{SYY/(n-1)} \end{aligned} \quad (٣٦-٣)$$

وبحل (٣٦-٣) نحصل على:

$$\bar{R}^2 = 1 - \left[ \frac{(1 - R^2)(n-1)}{n-k-1} \right] \quad (٣٧-٣)$$

ويمكن أن يصبح معامل التحديد المعدل أصغر عند إدخال متغير آخر إلى النموذج لأن النقص في  $SSE$  يمكن أن يكون أكبر من أن يعوض عن نقص درجة حرية واحدة في المقام  $n-k-1$  . يلاحظ الآتي في معامل التحديد المعدل:

• يأخذ قيمة أقل من قيم معامل التحديد غير المعدل.

- يمكن أن يأخذ قيمة سالبة في حين نجد أن قيم معامل التحديد غير المعدل تكون دائماً موجبة.

للمثال (٢-٣) فإن قيمة معامل التحديد  $R^2$  تصب من جدول (٣-٦) حيث:

$$R^2 = \frac{SSR}{SSY} = \frac{371246}{591818} = .627.$$

أي أن حوالي 0.627 من الاختلافات الموجودة في Y ترجع أسبابها إلى تأثير المتغيرين المستقلين  $x_1, x_2$ .

أما معامل التحديد المعدل فيحسب من الصيغة التالية:

$$\begin{aligned}\bar{R}^2 &= 1 - \frac{SSE/(n-k-1)}{SSY/(n-1)} \\ &= 1 - \frac{(220572)/8}{591818/10} \\ &= 0.534.\end{aligned}$$

أي أن 0.534 من التباين في Y ترجع إلى تأثير المتغيرين  $x_1, x_2$ .

كثيراً ما برامج الحاسب الآلي الخاصة بالإنحدار المتعدد تحسب كل من  $R^2, \bar{R}^2$ . في الفصل الخامس سوف يتضح أهمية  $R^2$  في إختيار أفضل المتغيرات في النموذج.

### (٣-١٢-٣) إختبارات تخص كل معامل الإنحدار

عادة يكون الإهتمام في إيجاد إختبارات فروض تخص كل معامل إنحدار جزئي. تلك الإختبارات تساعد في تقدير قيمة كل متغير مستقل في النموذج. على سبيل المثال فإن النموذج سوف يكون أكثر كفاءة بإضافة متغيرات مستقلة جديدة إلى النموذج أو إلى حذف واحد أو أكثر من المتغيرات الموجودة في النموذج. دائماً إضافة متغير لنموذج الإنحدار يؤدي إلى زيادة قيمة مجموع مربعات الإنحدار وإلى نقص مجموع المربعات للبوالي. عادة يؤدي إضافة متغير مستقل جديد إلى النموذج إلى زيادة التباين للقيمة المقدرة  $\hat{y}$  وعلى ذلك لابد أن نأخذ الحيلة ونضيف المتغيرات المستقلة التي لها قيمة حقيقية في تفسير الإستجابة. وأكثر من ذلك فإن إضافة متغير مستقل غير مهم سوف يزيد مجموع المربعات للبوالي والتي تقلل أهمية النموذج.

إختبار فرض العدم لمعنوية أي معامل إنحدار جزئي، ليكن  $\beta_i$  يكون على الشكل التالي :

$$H_0: \beta_i = 0$$

ضد الفرض البديل:

$$H_1: \beta_i \neq 0 .$$

عند قبول فرض العدم  $H_0: \beta_i = 0$  فهذا يعني أن المتغير المستقل  $x_i$  يمكن حذفه من النموذج. الإحصاء لهذا الاختبار هو:

$$T = \frac{B_i}{\sqrt{S^2 c_{ii}}} . \quad (٣٨-٣)$$

حيث  $c_{ii}$  هو العنصر رقم  $i$  على القطر الرئيسي للمصفوفة  $(X'X)^{-1}$  والمقابل لـ  $B_i$ .

نرفض فرض العدم  $H_0: \beta_i = 0$  إذا كانت  $|t| > t_{\alpha/2}(n-k-1)$  في الحقيقة

هذا الاختبار جزئي أو إختبار هامشي لأن معامل الإنحدار  $b_i$  يعتمد على كل المتغيرات المستقلة الأخرى  $x_{i'} (i \neq i')$  في النموذج. أي أنه إختبار لإسهام  $x_i$  إذا علم أن المتغيرات الأخرى موجودة أصلاً في النموذج. للمثال (٣-٢) فإن العنصر الرئيسي على القطر للمصفوفة  $(X'X)^{-1}$  والمقابل لـ  $b_1$  هي  $0.00922992 = c_{11}$  ,  $b_1 = 9.61221$  ,  $s^2 = 27571.5$  . وعلى ذلك قيمة  $t$  المحسوبة هي :

$$t = \frac{b_1}{\sqrt{s^2 c_{11}}} = \frac{9.61221}{\sqrt{(27571.5)(0.00922992)}} = 0.602551 .$$

وبما أن  $t_{0.025}(8) = 2.306$  فإننا نقبل  $H_0: \beta_1 = 0$  أي أن المتغير  $x_1$  غير معنوي في النموذج إذا علم أن  $x_2$  موجود أصلاً في النموذج .

لإختبار فرض العدم :

$$H_0: \beta_2 = 0$$

ضد الفرض البديل:

$$H_1: \beta_2 \neq 0 ,$$

وبما أن  $c_{22} = 0.000107453$  ,  $b_2 = 5.57653$  فإن قيمة  $t$  المحسوبة هي :

$$t = \frac{b_2}{\sqrt{s^2 c_{22}}} = \frac{5.57653}{\sqrt{(27571.5)(0.000107453)}} = 3.2398.$$

وبما أن  $t_{0.025}(8) = 2.306$  فإننا نرفض فرض العدم  $H_0: \beta_2 = 0$  أي أن المتغير  $x_2$  معنوي في النموذج إذا علم أن  $x_1$  موجود أصلاً في النموذج.

### (٣-١٢-٤) طريقة مجاميع المربعات الإضافية

#### Extra-sum-of-squares

أيضاً يمكن مباشرة تقدير إسهام مجموع المربعات لمتغير مستقل، لكن  $x_1$ ، إذا علم أن المتغيرات الأخرى  $x_{i'}$  حيث  $(i' \neq i)$  موجودة أصلاً في النموذج وذلك باستخدام طريقة مجموع المربعات الإضافية. وبصورة عامة يمكن استخدام هذه الطريقة لدراسة مساهمة فئة جزئية من المتغيرات المستقلة في النموذج. ليكن نموذج الإحداد بعدد  $k$  من المتغيرات المستقلة:

$$Y = X\beta + \varepsilon$$

حيث  $Y$  متجه بدرجة  $n \times 1$  و  $X$  مصفوفة بدرجة  $n \times p$  و  $\beta$  متجه من الدرجة  $p \times 1$  و  $\varepsilon$  متجه بدرجة  $n \times 1$  و  $p = k+1$ . بفرض أننا نرغب في تقدير ما إذا كانت فئة جزئية من  $r < k$  من المتغيرات المستقلة تساهم معنوياً في النموذج. بفرض أن متجه معاملات الإحداد يمكن تجزئته كالتالي:

$$\beta = \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix},$$

حيث  $\beta_1^*$  متجه بدرجة  $(p-r) \times 1$  و  $\beta_2^*$  متجه بدرجة  $r \times 1$ . الآن نرغب في اختبار الفرض:

$$H_0: \beta_2^* = 0$$

ضد الفرض البديل:

$$H_1: \beta_2^* \neq 0$$

يمكن كتابة النموذج كالتالي:

$$Y = X\beta + \varepsilon = X_1\beta_1^* + X_2\beta_2^* + \varepsilon,$$

حيث المصفوفة  $X_1$  بدرجة  $n \times (p-r)$  تمثل الأعمدة للمصفوفة  $X$  والتي ترتبط بالمتجه  $\beta_1^*$  والمصفوفة  $X_2$  بدرجة  $n \times r$  والتي تمثل الأعمدة للمصفوفة  $X$  التي

ترتبط بالمتجه  $\beta_2^*$  . هذا النموذج يسمى النموذج الكامل full model . للنموذج الكامل ، نعلم أن :

$$b = (X'X)^{-1} X'y .$$

مجموع مربعات الإنحدار لهذا النموذج سيكون :

$$SSR(\beta) = b'X'y$$

بدرجات حرية  $p$  وذلك بالاعتماد على جدول تحليل التباين المعطى في جدول (٣-٨) . وعلى ذلك متوسط مجموع المربعات للخطأ سوف يكون :

$$MSE = \frac{y'y - b'X'y}{n - p} .$$

لإيجاد مساهمة الحد  $\beta_2^*$  في النموذج ، نوفق النموذج المختزل :

$$Y = X_1\beta_1^* + \varepsilon .$$

مقدرات المربعات الصغرى للمعلمة  $\beta_1^*$  في النموذج المختزل ستكون :

$$b_1^* = (X_1'X_1)^{-1} X_1'y$$

مجموع مربعات الإنحدار ستكون :

$$SSR(\beta_1^*) = b_1^{*'} X_1'y .$$

بدرجات حرية  $p - r$  . مجموع مربعات الإنحدار الذي تعود إلى  $\beta_2^*$  إذا علم أن  $\beta_1^*$  موجودة أصلاً في النموذج سيكون :

$$SSR(\beta_2^* | \beta_1^*) = SSR(\beta) - SSR(\beta_1^*)$$

بدرجات حرية  $p - (p - r) = r$  . مجموع المربعات  $SSR(\beta_2^* | \beta_1^*)$  يسمى مجموع المربعات الإضافي والذي يعود إلى  $\beta_2^*$  وذلك لأنه يقيس الزيادة في مجموع مربعات الإنحدار الناتجة من إضافة المتغيرات  $x_k, x_{k-r+2}, \dots, x_{k-r+1}$  إلى النموذج الذي يحتوي أصلاً على المتغيرات  $x_1, x_2, \dots, x_{k-r}$  . الآن  $SSR(\beta_2^* | \beta_1^*)$  مستقل عن MSE وفرض العدم  $H_0: \beta_2^* = 0$  يختبر باستخدام الإحصاء :

$$F = \frac{SSR(\beta_2^* | \beta_1^*) / r}{MSE} \quad (٢٩-٣)$$



إذا كانت القيمة المحسوبة للإحصاء  $F$  تزيد عن القيمة الجدولية  $F_{\alpha}[r, n-p]$  عند مستوى معنوية  $\alpha$  نرفض فرض العدم ونستنتج أن واحد على الأقل من المعالم في  $\beta_2^*$  لا تساوي صفر وبالتالي على الأقل واحد من المتغيرات المستقلة  $x_{k-r+1}, x_{k-r+2}, \dots, x_k$  تساهم معنوياً في نموذج الانحدار . عادة يسمى الاختبار السابق باختبار  $F$  الجزئية partial F-test وذلك لأنه يقيس مساهمة المتغيرات المستقلة في  $X_2$  إذا علم أن المتغيرات الأخرى في  $X_1$  موجودة أصلاً في النموذج . لتوضيح هذه الطريقة بفرض النموذج التالي :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

مجاميع المربعات للانحدار سوف تكون :

$$SSR(\beta_1 | \beta_0, \beta_2, \beta_3) ,$$

$$SSR(\beta_2 | \beta_0, \beta_1, \beta_3) ,$$

$$SSR(\beta_3 | \beta_0, \beta_1, \beta_2) .$$

والتي لكل منها درجة حرية واحدة نقيس مساهمة كل متغير  $x_i$  حيث  $i=1,2,3$  في النموذج إذا علم أن كل المتغيرات الأخرى موجودة أصلاً في النموذج . أي أننا نقيس قيمة إضافة  $x_i$  إلى النموذج والذي لم يكن أصلاً موجود في النموذج . عموماً فإننا نوجد :

$$SSR(\beta_i | \beta_0, \beta_1, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_k) , \quad 1 \leq i \leq k$$

والتي تمثل الزيادة في مجموع مربعات الانحدار الذي يعود إلى إضافة  $x_i$  إلى النموذج الذي يحتوي أصلاً على  $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k$  . أي أنه مقياس لمساهمة  $x_i$  وكأنه آخر متغير يضاف إلى النموذج .

يمكن إثبات أن اختبار  $F$  الجزئي لمتغير مفرد  $x_i$  يكافئ اختبار  $t$  في (٣٨-٣) . في بعض الأحيان اختبار  $F$  الجزئي يعتبر طريقة عامة والتي بها يمكن قياس تأثير فئة من المتغيرات . في الفصل الخامس سوف نبين كيف أن اختبار  $F$  الجزئي يلعب دور رئيسي في بناء النموذج ، أي في البحث عن أفضل فئة من المتغيرات تستخدم في النموذج .

أيضاً يمكن استخدام طريقة مجموع المربعات الإضافية لاختبار أي فئة جزئية من المتغيرات المستقلة والتي تبدو مناسبة لمشكلة خاصة تحت التحليل . على سبيل المثال ليكن النموذج التالي من الدرجة الثانية :

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

وإذا كان الإهتمام في إيجاد  $SSR(\beta_1 | \beta_0)$  والذي يقيس التأثير الخطي لـ  $x$  و  $(\beta_2 | \beta_0, \beta_1)$  والذي يقيس مساهمة إضافة حد التربيع إلى النموذج المحتوي أملاً على حد الخطية . أي أن هذا الأسلوب ، بصورة عامة ، يساعد على متابعة اختبار مدى معنوية التأثير الذي يضيفه كل متغير مستقل على المتغير التابع وبالتالي إنتقاء العلاقة الدالة الملائمة بين المتغيرات المدروسة وعلى ذلك فإنه يمكن تجزئة مجموع مربعات الانحدار إلى مكونات هامشية بدرجة حرية واحدة على سبيل المثال التالي للنموذج:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

حيث :

$$SSY = SSR(\beta_1, \beta_2, \beta_3 | \beta_0) + SSE .$$

يمكننا تجزئة  $SSR(\beta_1, \beta_2, \beta_3 | \beta_0)$  بثلاثة درجات حرية كالتالي :

$$\begin{aligned} SSR(\beta_1, \beta_2, \beta_3 | \beta_0) &= SSR(\beta_1 | \beta_0) \\ &+ SSR(\beta_2 | \beta_1, \beta_0) + SSR(\beta_3 | \beta_1, \beta_2, \beta_0) \end{aligned}$$

حيث كل مجموع مربعات في الطرف الأيمن له درجة حرية واحدة. ويجب أن نعلم ترتيب المتغيرات في هذه المكونات إختياري. فعلى سبيل المثال يمكن تجزئة  $SSR(\beta_1, \beta_2, \beta_3 | \beta_0)$  كالتالي :

$$\begin{aligned} SSR(\beta_1, \beta_2, \beta_3 | \beta_0) &= SSR(\beta_2 | \beta_0) \\ &+ SSR(\beta_1 | \beta_2, \beta_0) + SSR(\beta_3 | \beta_1, \beta_2, \beta_0) \end{aligned}$$

في بعض الأحيان فإن طريقة مجموع المربعات ليست دائماً الطريقة لتجزئة مجموع مربعات الانحدار وذلك لأنه عموماً :

$$\begin{aligned} SSR(\beta_1, \beta_2, \beta_3 | \beta_0) &\neq SSR(\beta_1 | \beta_2, \beta_3, \beta_0) \\ &+ SSR(\beta_2 | \beta_1, \beta_3, \beta_0) \\ &+ SSR(\beta_3 | \beta_1, \beta_2, \beta_0) . \end{aligned}$$

ونظراً لأن معظم حزم الحاسب الآلي الخاصة بالانحدار تعتمد في حساب مجموع المربعات الإضافية على جدول تحصيل التباين في (٣-٦) لذلك سوف نستخدم الصيغ

الآتية في حساب مجموع المربعات الإضافية. مجموع مربعات الإنحدار للنموذج الكامل سوف تكون على الصورة الآتية:

$$SSR(\beta_1, \beta_2, \dots, \beta_{p-1} | \beta_0) = b'X'y - \frac{(\sum y_i)^2}{n}$$

بدرجات حرية  $p-1$  . متوسط مجموع المربعات للخطأ سوف يكون :

$$MSE = \frac{y'y - b'X'y}{n - p}$$

و بالنسبة للنموذج المختزل فإن مجموع المربعات الإنحدار سوف يكون :

$$SSR(\beta_1, \beta_2, \dots, \beta_{p-r-1} | \beta_0) = b'^*X_1'y - \frac{(\sum y_i)^2}{n}$$

بدرجات حرية  $p-r-1$ .

وعلى ذلك مجموع مربعات الإنحدار الذي يعود إلى  $\beta_2^*$  إذا علم أن  $\beta_1^*$  موجودة أصلاً في النموذج سيكون :

$$SSR(\beta_2^* | \beta_1^*) = SSR(\beta_1, \beta_2, \dots, \beta_{p-1} | \beta_0) - SSR(\beta_1, \beta_2, \dots, \beta_{p-r-1} | \beta_0).$$

سوف نوضح طريقة الحساب في المثال التالي:

### مثال (٣-٥)

رغب باحث في مؤسسة علمية في تثمين العلاقة بين الرواتب السنوية لباحثين في الرياضيات من مستوى متوسط ومتقدم (Y، بآلاف الدولارات ) ورقم قياسي يعبر عن نوعية المنشورات ( $x_1$ ) ، عدد سنوات الخبرة ( $x_2$ ) ، ورقم قياسي يعبر عن النجاح في الحصول على دعم منحة ( $x_3$ ) . يعطي جدول (٣-٩) البيانات لعينة من 24 باحثاً في الرياضيات من مستويات متوسطة ومتقدمة.

جدول (٣-٩)

$x_1$	$x_2$	$x_3$	$y$
3.5	9	6.1	33.2
5.3	20	6.4	40.3
5.1	18	7.4	38.7
5.8	33	6.7	46.8
4.2	31	7.5	41.4
6	13	5.9	37.5
6.8	25	6	39
5.5	30	4	40.7
3.1	5	5.8	30.1
7.2	47	8.3	52.9
4.5	25	3	38.2
4.9	11	6.4	31.8
8	23	7.6	43.3
6.5	35	7	44.1
6.6	39	5	42.8
3.7	21	4.4	33.6
6.2	7	5.5	34.2
7	40	7	48
4	35	6	38
4.5	23	3.5	35.9
5.9	33	4.9	40.4
5.6	27	4.3	36.8
4.8	34	8	45.2
3.9	15	5	35.1

وبفرض أن معادلة الانحدار الخطي المتعدد المقدرة هي:

$$\hat{y} = 17.8469 + 1.10313 x_1 + 0.32152 x_2 + 1.28894 x_3.$$

لإيجاد مدى مساهمة المتغير  $x_2$  في النموذج فإن الفرض المناسب سيكون :

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0 .$$

لإختبار هذا الفرض ، فلنحتاج إلى مجموع المربعات الإضافية والذي يعود إلى  $\beta_2$  أو :

$$\begin{aligned} SSR(\beta_2 | \beta_1, \beta_3, \beta_0) &= SSR(\beta_1, \beta_2, \beta_3, \beta_0) - SSR(\beta_1, \beta_3, \beta_0) \\ &= SSR(\beta_1, \beta_2, \beta_3 | \beta_0) - SSR(\beta_1, \beta_3 | \beta_0). \end{aligned}$$

بدرجة حرية واحدة. جدول تحليل التباين عندما  $x_1, x_2, x_3$  موجودين أصلاً في النموذج معطى في جدول (٣-١٠) وذلك حتى يمكننا استخدام جدول تحليل التباين على الصورة المعطاة في جدول (٣-٦).

جدول (٣-١٠)

S.O.V	df	SS	MS	F
regression	3	627.817	209.272	68.1192
residual	20	61.443	3.07215	—
Total	23	689.26	—	—

جدول تحليل التباين عندما  $x_1, x_3$  في النموذج معطى في جدول (٣-١١).

جدول (٣-١١)

S.O.V	df	SS	MS	F
regression	2	397.192	198.596	14.2792
residual	21	292.068	13.908	—
Total	23	689.26	—	—

يشير جدول (٣-١١) عند وجود  $x_1, x_2, x_3$  في نموذج الإنحدار إلى أن مجموع مربعات الإنحدار هو :

$$SSR(\beta_1, \beta_2, \beta_3 | \beta_0) = 627.817,$$

وأن مجموع مربعات الخطأ هو :

$$SSE(x_1, x_2, x_3) = 61.443.$$

ونلاحظ من جدول (٣-١١) أن مجموع مربعات الإنحدار عند وجود  $x_1, x_3$  في النموذج هي :

$$SSR(\beta_1, \beta_3 | \beta_0) = 397.192$$

وأن مجموع مربعات الخطأ هي :

$$SSE(x_1, x_3) = 292.068$$

ونلاحظ أن مجموع المربعات عند وجود  $(x_1, x_2, x_3)$  في النموذج  $SSE(x_1, x_2, x_3)$  أقل من قيمته عندما يتضمن النموذج  $x_1, x_3$  والفرق بينهما هو مجموع المربعات الإضافي وهو :

$$SSR(\beta_2 | \beta_1, \beta_3, \beta_0)$$

$$\begin{aligned}
 &= SSE(x_1, x_3) - SSE(x_1, x_2, x_3) \\
 &= 292.068 - 61.443 \\
 &= 230.625.
 \end{aligned}$$

وهذا التخفيض في مجموع مربعات الخطأ كان نتيجة لإضافة  $x_2$  إلى نموذج الإنحدار إذا علم أن  $x_1, x_3$  موجود أصلاً في النموذج. وهكذا يقيس مجموع المربعات الإضافي  $SSR(\beta_2 | \beta_1, \beta_3, \beta_0)$  التأثير الهامشي لإضافة  $x_2$  إلى نموذج إنحدار كان يقتصر على  $x_1, x_3$  ويعكس الرمز  $SSR(\beta_2 | \beta_1, \beta_3, \beta_0)$  هذا التخفيض الإضافي في مجموع مربعات الخطأ الذي يترافق مع  $x_2$  إذا علم أن  $x_1, x_3$  كانت أصلاً في النموذج. وقيس مجموع المربعات الإضافي  $SSR(\beta_2 | \beta_1, \beta_3, \beta_0)$  بصورة مكافئة، للزيادة الهامشية في مجموع مربعات الإنحدار حيث:

$$\begin{aligned}
 &SSR(\beta_2 | \beta_1, \beta_3, \beta_0) \\
 &= SSR(\beta_1, \beta_2, \beta_3 | \beta_0) \\
 &- SSR(\beta_1, \beta_3 | \beta_0) \\
 &= 397.192 - 627.817 \\
 &= 230.625.
 \end{aligned}$$

وسبب تكافؤ التخفيض الهامشي في مجموع مربعات الخطأ والزيادة الهامشية في مجموع مربعات الإنحدار هو أنه في متطابقة تحليل التباين نجد أن:

$$SYY = SSE + SSR.$$

وبما أن  $SYY$  يعتمد على المشاهدات  $y_i$  وبالتالي لا يعتمد على النموذج الذي جرى توقيسه فإن أي تخفيض في  $SSE$  يتضمن زيادة مطابقة في  $SSR$ . لإختبار  $H_0: \beta_2 = 0$  فإننا نحسب قيمة للإحصاء  $F$  حيث:

$$\begin{aligned}
 F &= \frac{SSR(\beta_2 | \beta_1, \beta_3, \beta_0) / 1}{MSE} \\
 &= \frac{230.625 / 1}{3.07215} = 75.0696.
 \end{aligned}$$

ويجب أن نتذكر أن  $MSE$  من النموذج الكامل باستخدام  $x_1, x_2, x_3$  يستخدم في المقام للإحصاء  $F$ . وبما أن قيمة  $F$  الجدولية  $F_{0.05[1,20]} = 4.35$  فإننا نرفض فرض العدم  $H_0: \beta_2 = 0$  ونستنتج أن المتغير  $x_2$  يساهم معنوياً في النموذج. وبما أن إختبار  $F$  الجزئي يشمل على متغير واحد، فهو يكافئ إختبار  $t$  وبما أن

مربع الإحصاء  $t$  بدرجات حرية  $v$  هو الإحصاء  $F$  بدرجات حرية  $(1, v)$   
 $t = \sqrt{75.0696} = 8.66427$

(٥-١٢-٣) إختبار فرضية حول أهمية تعاقب المتغيرات

في هذه الحالة ندخل المتغيرات المستقلة حسب الأهمية ونبعا لخبرة سابقة أن يكون المتغير  $x_1$  أهم المتغيرات يليه المتغير  $x_2$ ... وهكذا حتى المتغير  $x_k$  القليل الأهمية . ويكون الهدف من هذا الإختبار هو الإجابة على السؤال التالي : هل يمكن حذف  $x_k$  من النموذج ؟ وإذا كان الجواب بنعم فهل يمكن حذف  $x_{k-1}$  وهكذا . أي أن فروض العدم لهذا الإختبار سوف تكون على الشكل التالي :

$$H_{10} : \beta_1 = 0 , H_{11} : \beta_1 \neq 0$$

$$H_{20} : \beta_2 = 0 , H_{21} : \beta_2 \neq 0$$

$$\vdots \quad \quad \quad \vdots$$

$$H_{k0} : \beta_k = 0 , H_{k1} : \beta_k \neq 0 .$$

وفي هذا الإختبار نبدأ بإختبار هل  $H_{k0} : \beta_k = 0$  ضد الفرض البديل  $H_{k1} : \beta_k \neq 0$  . الإحصاء لهذا الإختبار يأخذ الشكل التالي :

$$F = \frac{MSR(\beta_k | \beta_1, \beta_2, \dots, \beta_{k-1}, \beta_0)}{MSE(\beta_1, \beta_2, \dots, \beta_k | \beta_0)}$$

فإذا قبلنا فرض العدم  $H_0$  فإننا نحذف الحد  $\beta_k x_k$  من النموذج ونبدأ في إختبار فرض العدم  $H_{k-1,0} : \beta_{k-1} = 0$  وباستخدام الإحصاء  $F$  الذي يأخذ الشكل التالي .:

$$F = \frac{MSR(\beta_{k-1} | \beta_1, \beta_2, \dots, \beta_{k-2}, \beta_0)}{MSE(\beta_1, \beta_2, \dots, \beta_{k-1} | \beta_0)}$$

وهكذا ..

جدول تحليل التباين معطى في جدول (١٢-٣) .

جدول (٢-١٢)

S.O.V	df	SS
$\beta_1, \beta_2, \dots, \beta_k   \beta_0$	k	$SSR(\beta_1, \beta_2, \dots, \beta_k   \beta_0)$
$\beta_1   \beta_0$	1	$SSR(\beta_1   \beta_0)$
$\beta_2   \beta_1, \beta_0$	1	$SSR(\beta_2   \beta_1, \beta_0)$
$\beta_3   \beta_1, \beta_2, \beta_0$	1	$SSR(\beta_3   \beta_1, \beta_2, \beta_0)$
$\vdots$	$\vdots$	$\vdots$
$\beta_k   \beta_1, \dots, \beta_{k-1}$	1	$SSR(\beta_k   \beta_1, \beta_2, \dots, \beta_{k-1}, \beta_0)$
الخطأ	$n - k - 1$	$SSE(\beta_1, \beta_2, \dots, \beta_k, \beta_0)$
الكلية	$n - 1$	

ويوجد الكثير من حزم الحاسب الآلي الخاصة بالإحصاء لتجزئة SSR إلى مجاميع مربعات إضافية كل منها بدرجة حرية واحدة . ويكون ذلك بالترتيب نفسه الذي أدخلت فيه المتغيرات المستقلة إلى النموذج . فعلى سبيل المثال في وجود ثلاثة متغيرات مستقلة  $x_1, x_2, x_3$  وإذا أدخلت بالترتيب  $x_3, x_2, x_1$  فإن مجاميع المربعات الإضافية المعطاة في المخرجات سوف تكون :

$$SSR(\beta_1 | \beta_0)$$

$$SSR(\beta_2 | \beta_1, \beta_0)$$

$$SSR(\beta_3 | \beta_1, \beta_2, \beta_0).$$

وعند الرغبة بمجموع مربعات إضافية يحتوي على عدة متغيرات مستقلة فيمكن الحصول عليه بجمع ما يناسب من مجاميع للمربعات الإضافية بدرجة واحدة من الحرية . فعلى سبيل المثال يمكن الحصول على  $R(\beta_2, \beta_3 | \beta_1, \beta_0)$  للمثال السابق بجمع  $SSR(\beta_2 | \beta_1, \beta_0)$  و  $SSR(\beta_3 | \beta_1, \beta_2, \beta_0)$  . وعند الرغبة في الحصول على مجموع المربعات الإضافية  $SSR(\beta_1, \beta_3 | \beta_2, \beta_0)$  من حزم حاسب آلي تقدم مجاميع مربعات إضافية بدرجة حرية واحدة بالترتيب الذي أدخلت فيه المتغيرات المستقلة فسوف نحتاج إلى أن تكون المتغيرات المستقلة قد أدخلت بالترتيب  $x_2$  ثم  $x_1$  ثم  $x_3$  أو  $x_2$  ثم  $x_3$  ثم  $x_1$  والترتيب الأول يعطي:



$$SSR(\beta_2|\beta_0)$$

$$SSR(\beta_1|\beta_2, \beta_0)$$

$$SSR(\beta_3|\beta_1, \beta_2, \beta_0)$$

ومجموع مجموعي المربعات الإضافيين الآخرين سيعطي  $SSR(\beta_1, \beta_3|\beta_2, \beta_0)$

مثال (٦-٣)

للبانات الموجودة في جدول تحليل التباين المعطى في جدول (١٣-٣) والخاص بالاختبار التتابعي :

لختبر فرض العدم

$$H_0 : \beta_4 = 0$$

ضد الفرض البديل:

$$H_1 : \beta_4 \neq 0$$

جدول (١٣-٣)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2, \beta_3, \beta_4   \beta_0$	4	3429.273		
$\beta_1   \beta_0$	1	216.256	216.256	
$\beta_2   \beta_1, \beta_0$	1	309.851	309.851	
$\beta_3   \beta_1, \beta_2, \beta_0$	1	29.214	29.214	
$\beta_4   \beta_1, \beta_2, \beta_3, \beta_0$	1	2873.952	2873.952	575.63
الخطأ	27	134.804	4.993	
الكلي	31			

الحل

نحسب قيمة F من الصيغة التالية :

$$F = \frac{SSR(\beta_4 | \beta_1, \beta_2, \beta_3, \beta_0)}{SSE(\beta_1, \beta_2, \beta_3, \beta_4 | \beta_0)} \cdot \frac{1}{n-k-1}$$

$$= \frac{2873.952 / 1}{134.804 / 27} = 575.63 .$$

وبما أن قيمة F المحسوبة ( 575. 63 ) تزيد عن القيمة الجدولية  $F_{0.05}[1,27] = 4.21$  فهذا يعني أن إضافة  $x_4$  إلى معادلة الانحدار المحتوية على  $x_1, x_2, x_3$  يساعد في التنبؤ بـ Y .

(٣-١٢-٦) الحالة الخاصة لأعمدة متعامدة في المصفوفة X

بفرض أن المصفوفة X لنموذج الانحدار المتعدد على الشكل :

$$X = [I, X_1, X_2, \dots, X_k]$$

حيث  $I$  يمثل العمود الذي عناصره كلها الواحد الصحيح و  $X_j$  متجه عمود يمثل مستويات  $x_j$  . عندها :

$$X'_m X_\ell = 0, \quad m \neq \ell ,$$

يقال للمتجهين  $X_\ell$  و  $X_m$  أنهما متعامدين لبعضهما . في حالة التعامد الكامل حيث  $X'_m X_\ell = 0$  لكل قيم  $m$  و  $\ell$  و  $m = 1, 2, \dots, k$  و  $\ell = 1, 2, \dots, k$  فإن:

$$\sum_{j=1}^n x_{ij} = 0 \quad , \quad i = 1, 2, \dots, k .$$

عندما تكون الأعمدة في X متعامدة فإن المصفوفة  $X'X$  تكون مصفوفة قطرية وعلى ذلك المعادلات الطبيعية لنموذج الانحدار الخطي المتعدد تختزل إلى :

$$nb_0 = \sum y_j ,$$

$$b_1 \sum_{j=1}^n x_{1j}^2 = \sum_{j=1}^n x_{1j} y_j ,$$

$$\vdots \quad \quad \quad \vdots$$

$$b_k \sum_{j=1}^n x_{kj}^2 = \sum_{j=1}^n x_{kj} y_j .$$

كمثال لنموذج إحصائي بمتغيرات متعامدة، بفرض النموذج  
 $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$  حيث  $Y$  هي المصفوفة  $X$  على الشكل التالي :

$$X = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

فإن مستويات المتغيرات تقابل تصميم عاملي  $2^3$  . من السهولة ملاحظة أن الأعمدة في  $X$  متعامدة وعلى ذلك  $SSR(\beta_i), i=1,2,3$  بقيس مساهمة المتغير المستقل  $x_i$  في النموذج بصرف النظر عن وجود أو عدم وجود المتغيرات المستقلة الأخرى في النموذج .

واحد من المميزات هنا هو سهولة تجزئة  $SSR$  إلى مكونات وكل مكون بدرجة حرية واحدة .

في حالة التعماد يمكن كتابة :

$$\begin{aligned} SSR &= \sum (\hat{y}_j - \bar{y})^2 = \sum (b_0 + b_1 x_{1j} + \dots + b_k x_{kj} - b_0)^2 \\ &= b_1^2 \sum_{j=1}^n x_{1j}^2 + b_2^2 \sum_{j=1}^n x_{2j}^2 + \dots + b_k^2 \sum_{j=1}^n x_{kj}^2 \\ &= R(\beta_1) + R(\beta_2) + \dots + R(\beta_k) . \end{aligned}$$

الرمز  $R(\beta_i)$  يمثل الكمية من مجموع مربعات الانحدار التي ترتبط بنموذج يحتوي على متغير مستقل  $x_i$  . لإختبار معنوية فئة من المتغيرات في أن واحد في حالة التعماد فإن مجموع المربعات الانحدار تصبح :

$$R(\beta_1, \beta_2, \dots, \beta_r | \beta_{r+1}, \dots, \beta_k, \beta_0) \\ = R(\beta_1 | \beta_0) + R(\beta_2 | \beta_0) + \dots + R(\beta_r | \beta_0) .$$

حيث  $r < k$ .

وحالة خاصة منها هي :

$$R(\beta_1 | \beta_1, \beta_2, \dots, \beta_k, \beta_0) = R(\beta_1 | \beta_0) .$$

وذلك في حالة تقويم متغير مستقل مفرد. يعطى جدول (٣-١٤) التباين الكلي للاستجابة والذي يجرىء إلى مكونات وكل مكون بدرجة حرية وهذا بالإضافة إلى حد الخطأ بدرجة حرية  $n - p$ .

فرض العدم سيكون :

$$H_0 : \beta_i = 0$$

ضد الفرض البديل :

$$H_1 : \beta_i \neq 0 \quad \text{و} \quad i = 1, 2, \dots, k .$$

الإحصاء  $F$  سوف يكون :

$$F \approx \frac{R(\beta_i | \beta_0)}{MSE} .$$

إذا زادت قيمة  $F$  المحسوبة عن قيمة  $F$  الجدولية عند مستوى معنوية  $\alpha$  بدرجات حرية  $n - p, 1$  فإننا نرفض فرض العدم.

جدول (١٤-٣)

S.O.V	df	SS	MS
$\beta_1$	1	$R(\beta_1   \beta_0) = b_1^2 \sum_{j=1}^n x_{1j}^2$	$R(\beta_1   \beta_0)$
$\beta_2$	1	$R(\beta_2   \beta_0) = b_2^2 \sum_{j=1}^n x_{2j}^2$	$R(\beta_2   \beta_0)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\beta_k$	1	$R(\beta_k   \beta_0) = b_k^2 \sum_{j=1}^n x_{kj}^2$	$R(\beta_k   \beta_0)$
الخطأ	n-p	SSE	$s^2 = \frac{SSE}{n-p} = MSE$
الكلية	n-1		

مثال (٧-٣)

بفرض تجربة لدراسة تأثير متغير الإستجابة Y على ثلاثة متغيرات مستقلة والبيانات معطاة في جدول (١٥-٣)

جدول (١٥-٣)

y	$x_1^*$	$x_2^*$	$x_3^*$
82	150(-1)	12(-1)	220(-1)
93	190(1)	12(-1)	220(-1)
114	150(-1)	24(1)	220(-1)
124	150(-1)	12(-1)	250(1)
111	190(1)	24(1)	220(-1)
129	190(1)	12(-1)	250(1)
157	150(-1)	24(1)	250(1)
164	190(1)	24(1)	250(1)

البيانات في جدول (١٥-٣) لتجربة في عاملة  $2 \times 2 \times 2$  ( $2^3$ ) في تصميم التجارب، لوجد نموذج الإتحدار المتعدد المقدر وقدر تأثير كل عامل في النموذج .

### الحل

يلاحظ أن كل متغير له مستويين . وقد تم تحويل البيانات على المتغيرات المستقلة  $x_i$  كشجرة لتسهيل الحساب وذلك تبعاً للصيغ الآتية :

$$x_1 = \frac{x_1^* - 170}{20} ,$$

$$x_2 = \frac{x_2^* - 18}{6} ,$$

$$x_3 = \frac{x_3^* - 235}{15} .$$

المستويات الناتجة لكل  $x_1, x_2, x_3$  تأخذ القيم -1 أو +1 كما هو موضح في جدول (١٥-٣) . المصفوفة  $X$  تصبح :

$$X = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

حيث الأعمدة في المصفوفة  $X$  متعامدة . وعلى ذلك يمكن حساب معاملات نموذج الإتحدار كالتالي :

$$b_0 = \frac{\sum_{j=1}^n y_j}{8} = 121.75,$$

$$b_1 = \frac{\sum_{j=1}^n x_{1j} y_j}{\sum x_{1j}^2} = \frac{20}{8} = 2.5,$$

$$b_2 = \frac{\sum_{j=1}^n x_{2j} y_j}{\sum x_{2j}^2} = \frac{118}{8} = 14.75 ,$$

$$b_3 = \frac{\sum_{j=1}^n x_{3j} y_j}{\sum x_{3j}^2} = \frac{174}{8} = 21.75.$$

وعلى ذلك معادلة الإتحاد المقدرة هي :

$$\hat{y} = 121.75 + 2.5x_1 + 14.75x_2 + 21.75x_3 .$$

يوضح جدول تحليل التباين في جدول (١٦-٣) SSR لكل متغير . عند مقارنة قيمة F المحسوبة لكل متغير مع قيمة F الجدولية عند  $\alpha = 0.05$  ودرجات حرية 1,4 (F(1,4) = 7.71) يتضح أن  $x_1$  غير معنوي عند  $\alpha = 0.05$  بينما المتغيرين .05

$x_2, x_3$  معنويين.

جدول (١٦-٣)

S.O.V	df	SS	MS	F
$\beta_1$	1	$(2.5)^2(8) = 50$	50	2.16
$\beta_2$	1	$(14.75)^2(8) = 1740.50$	1740.50	75.26*
$\beta_3$	1	$(214.75)^2(8) = 3784.50$	3784.50	163.65*
الخطأ	4	92.5	23.1250	
الكل	7	5667.50		

(٧-١٢-٣) إختبار الفرض الخطي العام  $TB=0$

في بعض الأحيان قد نفترض أو نقترح نماذج أكثر عمومية مما نحتاجه للتعبير عن العلاقة بين متغير الإستجابة مع المتغيرات المستقلة المقترحة في العلاقة . بفرض أن باحث قام بدراسة العلاقة بين متغير الإستجابة  $Y$  مع متغيرين مستقلين  $x_1, x_2$  فإذا كان كلا المتغيرين ضروريان في النموذج فإن النموذج:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

يكون نموذجاً صحيحاً . أما إذا كان الباحث يتوقع أن التأثير الأفضل هو الفرق  $x_1 - x_2$  فإن النموذج :

$$Y = \beta_0 + \beta^* (x_1 - x_2) + \varepsilon$$

يكون نموذجاً ملائماً جداً . السؤال الآن هو أي من النماذج هو الأفضل وكيف يمكن التحقق من ذلك حيث يمكن الإجابة على هذا السؤال من خلال الإجابة على الافتراض التالي:

$$\beta_1 = -\beta_2$$

أو

$$\beta_1 + \beta_2 = 0$$

أي من خلال إختبار الفرضية التالية :

$$H_0 : \beta_1 + \beta_2 = 0$$

ضد الفرض البديل:

$$H_1 : \beta_1 + \beta_2 \neq 0 .$$

عند قبول فرض العدم نقبل النموذج الأول وعند رفض فرض العدم نقبل النموذج الثاني . بما أن الفرضية  $H_0$  تتضمن على تركيبة خطية بالنسبة للمعالم  $\beta_1, \beta_2$  لذلك نسمى فرضية خطية. الفرضية الخطية يمكن أن تحتوي على أكثر من معادلة أو علاقة واحدة حول المعالم وكمثال آخر بفرض نموذج الإنحدار هو:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

حيث :

$$H_0 : \beta_1 + \beta_2 = 0$$

$$\beta_2 - \beta_3 = 0 ,$$

$$\beta_1 + 3\beta_2 - 2\beta_3 = 0$$



في هذا المثال لدينا فرضية خطية فيها علاقتين خطيتين مستقلتين فقط والسبب لأن العلاقة الثالثة  $0 = \beta_1 + 3\beta_2 - 2\beta_3$  عبارة عن تركيبة خطية للعلاقتين الأولى والثانية كما يلي :

$$\beta_1 + 3\beta_2 - 2\beta_3 = 1(\beta_1 + \beta_2) + 2(\beta_2 - \beta_3).$$

وكمثال لفرضية خطية كل العلاقات بها مستقلة إذا كان نموذج الإتحاد المقدر على الشكل :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$$

حيث :

$$H_0 : \beta_1 + \beta_2 = 0, \beta_2 - \beta_3 = 0,$$

$$\beta_3 - \beta_4 = 0, \beta_4 - \beta_5 = 0$$

هنا  $H_0$  فرضية خطية فيها أربع علاقات خطية مستقلة.

بفرض أن الفرضية الخطية موضع الإهتمام يمكن التعبير عنها كالتالي :

$$H_0 : TB = 0$$

حيث  $T$  مصفوفة من الدرجة  $m \times p$  من الثوابت ، بحيث أن  $r$  فقط من  $m$  من المعادلات في  $TB = 0$  مستقلة . النموذج الكامل هو  $Y = X\beta + \varepsilon$  حيث  $b = (X'X)^{-1}X'y$  و مجموع مربعات الخطأ (البواقي) للنموذج الكامل هو :

$$SSE(FM) = y'y - b'X'y,$$

بدرجات حرية  $n - p$  . للحصول على النموذج المختزل ، حيث  $r$  من المعادلات في  $TB = 0$  مستقلة والتي تستخدم في حل  $r$  من معاملات الإتحاد في النموذج الكامل بدلالة معاملات الإتحاد الباقية والتي عددها  $p - r$  . وهذا يؤدي إلى النموذج المختزل  $Y = Z\gamma + \varepsilon$  ، على سبيل المثال  $Z$  مصفوفة من الدرجة  $n \times (p - r)$  و  $\gamma$  متجه من الدرجة  $(p - r) \times 1$  من معاملات الإتحاد المجهولة . التقدير لها هو :

$$\hat{\gamma} = (Z'Z)^{-1}Z'y$$

ومجموع مربعات الخطأ للنموذج المختزل هو :

$$SSE(RM) = y'y - \hat{\gamma}'Z'y$$

بدرجات حرية  $n - p + r$  . إن النموذج المختزل يحتوي على معاملات أقل من النموذج الكامل وعلى ذلك  $SSE(RM) \geq SSE(FM)$  . لإختبار الفرض  $H_0 : TB = 0$  فإننا نستخدم الفرق في مجاميع مربعات البواقي :

$$SSH = SSE(RM) - SSE(FM) .$$

بدرجات حرية  $n-p+r-(n-p)=r$  . هنا SSH يسمى مجموع المربعات الذي يعود إلى الفرض  $H_0:TB=0$  . الإحصاء المناسب للاختبار هو :

$$F = \frac{SSH / r}{SSE(FM) / (n-p)} . \quad (٤٠-٣)$$

سوف نرفض  $H_0:TB=0$  إذا كانت قيمة  $F$  المحسوبة من الإحصاء  $F$  تزيد عن القيمة الجدولية  $F_{\alpha}[r, n-p]$  .  
في الجزء الثاني سوف نقدم بعض الأمثلة للتوضيح.

#### اختبار تساوي معاملي الحدار

بفرض النموذج :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

للمodel الكامل فان  $SSE(FM)$  له  $n-p=n-4$  درجات حرية . سوف نختبر الفرض  $H_0: \beta_1 = \beta_3$  . هذا الفرض يمكن التعبير عنه كالتالي :

$$H_0:TB=0$$

حيث :

$$T = [0, 1, 0, -1]$$

هي متجه صف من الدرجة  $1 \times 4$  . في الحقيقة يوجد معادلة واحدة في  $TB=0$  ، تسمى  $\beta_1 - \beta_3 = 0$  . بالتعويض عن هذه المعادلة في النموذج الكامل فلنأخذ نحصل على النموذج المختزل :

$$\begin{aligned} Y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \\ &= \beta_0 + \beta_1 (x_1 + x_3) + \beta_2 x_2 + \varepsilon \\ &= \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \varepsilon \end{aligned}$$

حيث :

$$\gamma_0 = \beta_0 \text{ و } \gamma_1 = \beta_1 \text{ و } z_1 = x_1 + x_3 \text{ و } \gamma_2 = \beta_2 \text{ و } z_2 = x_2$$

سوف نحصل على  $SSE(RM)$  بدرجات حرية  $n-4+1=n-3$  عند توفيق النموذج المختزل .

مجموع المربعات الذي يعود إلى الفرض هو  $SSH = SS(RM) - SS(FM)$  بدرجات حرية  $1 = (n-4) - (n-3)$  . يمكن إختبار الفرض وذلك بحساب قيمة  $t$  حيث :

$$t = \frac{b_1 - b_3}{\sqrt{s^2(c_{11} + c_{33} - 2c_{13})}}$$

**إختبار  $TB = 0$**

بفرض النموذج :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

وإذا كان المطلوب إختبار فرض العدم  $H_0: \beta_1 = \beta_3, \beta_2 = 0$  ليكن :

$$T = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

سوف يكون هناك معادلتين في  $TB = 0$  وهي  $\beta_1 - \beta_3 = 0$  و  $\beta_2 = 0$  . هاتين المعادلتين تؤدي إلى النموذج المختزل :

$$Y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \varepsilon$$

$$= \beta_0 + \beta_1 (x_1 + x_3) + \varepsilon$$

$$= \gamma_0 + \gamma_1 z_1 + \varepsilon$$

في هذا المثال ،  $SSE(RM)$  له  $n - 2$  درجات حرية وعلى ذلك  $SSH$  له  $2 = (n - 4) - (n - 2)$  درجات حرية . الإحصاء لهذا الإختبار يأخذ الشكل التالي :

$$F = \frac{(SSH/2)}{SSE(FM)/(n-4)} . \quad (٤١-٣)$$

كمثال آخر بفرض أن لدينا نموذج الإنحدار الخطي التالي :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

الآن بإفتراض أننا نرغب في إختبار الفرضية الخطية التالية :

$$H_0: TB = 0$$

حيث :

$$T = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

أي أن فرض العدم  $H_0$  يمكن كتابته على الصورة التالية :

$$H_0 : \beta_1 - \beta_2 = 0$$

$$\beta_3 - \beta_4 = 0$$

ضد الفرض البديل :

$$H_1 : \beta_1 - \beta_2 \neq 0$$

$$\beta_3 - \beta_4 \neq 0.$$

أي أنه تحت فرض العدم فإن  $\beta_1 = \beta_2$  و  $\beta_3 = \beta_4$  . لحساب مجموع مربعات الخطأ من النموذج الكامل  $SSE(FM)$  بالتعويض في النموذج الأصلي عن  $\beta_2$  —  $\beta_1$  وعن  $\beta_4$  —  $\beta_3$  فإن النموذج الكامل يصبح :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$$

$$= \beta_0 + \beta_1 (x_1 + x_2) + \beta_3 (x_3 + x_4) + \varepsilon$$

وبوضع  $\beta_0 = \gamma_0$  و  $\beta_1 = \gamma_1$  و  $\beta_3 = \gamma_2$  و  $\beta_2 = \gamma_1$  و  $\beta_4 = \gamma_2$  و  $z_1 = x_1 + x_2$  و  $z_2 = x_3 + x_4$  فإن النموذج سيكون :

$$Y = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \varepsilon .$$

نحسب مجموع مربعات الخطأ من النموذج السابق ،  $SSE(RM)$  ، ومنها نحسب مجموع مربعات الخطأ للفرض  $SSH$  حيث :

$$SSH = SSE(RM) - SSE(FM) .$$

نحسب قيمة للإحصاء  $F$  من الصيغة التالية :

$$F = \frac{SSH / 2}{SSR(FM) / (n - p)}$$

جدول تحليل التباين موضح في جدول (٣-١٧)

جدول (٣-١٧)

S.O.V	df
$\beta_1, \beta_2, \beta_3, \beta_4   \beta_0$	4
$\gamma_1, \gamma_2   \gamma_0$	2
H الخطأ	2
الكل	

يتضح أن مفهوم مجموع المربعات الإضافي هو حالة خاصة لهذه الطريقة .

مثال (٣-٨)

للمثال (٣-٥) فإن :

نموذج الإنحدار الخطي المتعدد هو :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

الآن :

$$X'X = \begin{bmatrix} 24 & 128.6 & 599 & 143.7 \\ 128.6 & 727.44 & 3365.3 & 782.49 \\ 599 & 3365.3 & 17847 & 3671.9 \\ 143.7 & 782.49 & 3671.9 & 899.49 \end{bmatrix},$$

$$X'y = \begin{bmatrix} 948 \\ 5188.17 \\ 24873.7 \\ 5767.77 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 1.30446 & -0.101874 & 0.000442008 & -0.121579 \\ -0.101874 & 0.0353559 & -0.00167433 & -0.00764701 \\ 0.000442008 & -0.00167433 & 0.000448237 & -0.000443861 \\ -0.121579 & -0.00764701 & -0.000443861 & 0.0289992 \end{bmatrix}$$

-٢٠٠-

$$b = (X'X)^{-1}X'y = \begin{bmatrix} 17.8469 \\ 1.10313 \\ 0.32152 \\ 1.28894 \end{bmatrix},$$

$$\begin{aligned} SYY &= \sum y_j^2 - \frac{(\sum y_j)^2}{n} \\ &= 38135.3 - \frac{(948)^2}{24} \\ &= 689.26, \end{aligned}$$

$$\begin{aligned} SSR(\beta_1, \beta_2, \beta_3 | \beta_0) &= b'X'y - \frac{(\sum y_j)^2}{n} \\ &= 38073.8 - 37446 \\ &= 627.817, \end{aligned}$$

$$\begin{aligned} SSE(FM) &= SYY - SSR(\beta_1, \beta_2, \beta_3 | \beta_0) \\ &= 689.26 - 627.817 \\ &= 61.443. \end{aligned}$$

جدول تحليل التباين معطى في جدول (١٨-٣).

جدول (١٨-٣)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2, \beta_3   \beta_0$	3	627.817	209.272	68.1192
الخطأ	20	61.443	3.07215	
		689.26		
الكلي	23			

الآن لإختبار فرض العدم

$$H_0 : TB = 0 = \begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = 0.$$

أي أن :

$$H_0: \beta_2 - \beta_3 = 0$$

وعلى ذلك تحت فرض العدم

$$\beta_2 = \beta_3$$

ومن ثم نعوض عن  $\beta_3$  بـ  $\beta_2$  في النموذج الكامل :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

$$= \beta_0 + \beta_1 x_1 + \beta_2 (x_2 + x_3) + \varepsilon$$

فإذا رمزنا لـ

$$\beta_0 = \gamma_0$$

$$\beta_1 = \gamma_1$$

$$\beta_2 = \gamma_2$$

$$x_2 + x_3 = z_2, x_1 = z_1$$

فإن النموذج الكامل تحت فرض العدم يصبح كالآتي :

$$Y = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \varepsilon.$$

وعلى ذلك للمثال (٣-٥) فإن البيانات تصبح كما هو معطى في جدول (٣-١٩).

جدول (٣-١٩)

$z_1$	$z_2$	$y$
3.5	15.1	33.2
5.3	26.4	40.3
5.1	25.4	38.7
5.8	39.7	46.8
4.2	38.5	41.4
6.	18.9	37.5
6.8	31.	39.
5.5	34.	40.7
3.1	10.8	30.1
7.2	55.3	52.9
4.5	30.	38.2
4.9	17.4	31.8
8.	30.6	43.3
6.5	42.	44.1
6.6	44.	42.8
3.7	25.4	33.6
6.2	12.5	34.2
7.	47.	48.
4.	41.	38.
4.5	26.5	35.9
5.9	37.9	40.4
5.6	31.3	36.8
4.8	42.	45.2
3.9	20.	35.1

ومنها :

-۲.۲-

$$Z = \begin{bmatrix} 1 & 3.5 & 15.1 \\ 1 & 5.3 & 26.4 \\ 1 & 5.1 & 25.4 \\ 1 & 5.8 & 39.7 \\ 1 & 4.2 & 38.5 \\ 1 & 6 & 18.9 \\ 1 & 6.8 & 31. \\ 1 & 5.5 & 34. \\ 1 & 3.1 & 10.8 \\ 1 & 7.2 & 55.3 \\ 1 & 4.5 & 30. \\ 1 & 4.9 & 17.4 \\ 1 & 8 & 30.6. \\ 1 & 6.5 & 42. \\ 1 & 6.6 & 44. \\ 1 & 3.7 & 25.4 \\ 1 & 6.2 & 12.5 \\ 1 & 7 & 47. \\ 1 & 4 & 41. \\ 1 & 4.5 & 26.5 \\ 1 & 5.9 & 37.9 \\ 1 & 5.6 & 31.3 \\ 1 & 4.8 & 42. \\ 1 & 3.9 & 20. \end{bmatrix}$$

$$Z'Z = \begin{bmatrix} 24 & 128.6 & 742.7 \\ 128.6 & 727.44 & 4147.79 \\ 742.7 & 4147.79 & 26090.3 \end{bmatrix}$$



$$Z'y = \begin{bmatrix} 948 \\ 5188.17 \\ 30641.5 \end{bmatrix},$$

$$\gamma = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} 21.7383 \\ 1.29361 \\ 0.34997 \end{bmatrix}$$

جدول تحليل التباين معطى في جدول (٢٠-٢) .

جدول (٢٠-٢)

anova				
source	df	SS	MS	F
regression	2	596.965	298.482	67.914
residual	21	92.2952	4.39501	--
Total	23	689.26	--	--

وعلى ذلك :

$$SSE(RM) = 92.2952$$

$$SSE(FM) = 61.443$$

$$SSH = SSE(RM) - SSE(FM)$$

$$= 92.2952 - 61.443$$

$$= 30.8522,$$

$$F = \frac{SSH/1}{MSE} = \frac{30.8522}{3.07215} = 10.0425 .$$

وبما أن قيمة F المحسوبة تزيد عن القيمة الجدولية  $F_{0.05}[1,20] = 4.35$  فإننا نرفض فرض العدم:

$$H_0 : \beta_2 - \beta_3 = 0 .$$

أي نرفض النموذج:

$$Y = \beta_0 + \beta_1 + \beta_2(x_2 + x_3) + \varepsilon .$$

### (١٣-٣) معاملات الانحدار القياسية

#### Standardized Regression Coefficients

في بعض الأحيان يصعب مقارنة قيم معاملات نموذج الانحدار للمتغيرات المستقلة الموجودة في النموذج وذلك لإختلاف وحدات القياس للمتغيرات المستقلة. ولذلك فإن قيمة المعامل لامتد الباحث بمقياس يوضح أهمية المعامل في النموذج. ولذلك من الممكن التخلص من تأثير وحدات القياس المختلفة للمتغيرات على قيم معاملاتها وذلك بتحويلها إلى متغيرات معيارية ثم تقدير معالم نموذج الانحدار باستخدام طريقة المربعات الصغرى العادية. ويتم ذلك بأسلوبين :

الأسلوب الأول:

عن طريق حساب القيم التالية:

$$z_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i} , i = 1, 2, \dots, k , j = 1, 2, \dots, n \quad (٤٢-٣)$$

و

$$y_j^* = \frac{y_j - \bar{y}}{s_y} , j = 1, 2, \dots, n \quad (٤٣-٣)$$

حيث:

$$s_i^2 = \frac{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}{n-1}$$

هو تباين العينة للمتغير  $x_i$  و

$$s_y^2 = \frac{\sum_{j=1}^n (y_j - \bar{y})^2}{n-1} \quad (٤٤-٣)$$

هو تباين العينة للمتغير  $Y$  . المتغيرات  $Y$  في (٤٢-٣) و (٤٣-٣) لها متوسط عينة يساوي صفر وتباين عينة يساوي الواحد الصحيح. باستخدام تلك المتغيرات الجديدة فإن نموذج الانحدار يصبح:

$$Y_j^* = \beta_1^* z_{1j} + \beta_2^* z_{2j} + \dots + \beta_k^* z_{kj} + \epsilon_j , i = 1, 2, \dots, n \quad (٤٥-٣)$$

نلاحظ عدم وجود الجزء المقطوع في النموذج (٤٥-٣) . (في الحقيقة فإن تقدير  $\beta_0^* = \bar{y} = 0$  . مقدر المربعات الصغرى للمعالم في (٤٥-٣) هو:

$$b^* = (Z'Z)^{-1} Z'y^* \quad (٤٦-٣)$$

الأسلوب الثاني:

عن طريق حساب القيم التالية :

$$w_{ij} = \frac{x_{ij} - \bar{x}_i}{S_{ii}^{1/2}} \quad , j = 1, 2, \dots, n \quad , \quad i = 1, 2, \dots, k$$

و

$$y_j^0 = \frac{y_j - \bar{y}}{S_{YY}^{1/2}} \quad , j = 1, 2, \dots, n$$

حيث:

$$S_{ii} = \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$$

هو مجموع المربعات المصحح للمتغير المستقل  $x_i$ . كل متغير  $w_i$  له متوسط  $\bar{w}_i = 0$  وطول  $\sqrt{\sum_{j=1}^n (w_{ij} - \bar{w}_i)^2} = 1$  . بدلالة تلك المتغيرات فإن نموذج الإحداد يصبح:

$$Y_j^0 = \beta_1^* w_{1j} + \beta_2^* w_{2j} + \dots + \beta_k^* w_{kj} + \epsilon_j \quad , j = 1, 2, \dots, n$$

متجه معاملات المربعات الصغرى يصبح:

$$b^* = (W'W)^{-1} W'y^0 \quad (٤٧-٣)$$

المصفوفة  $W'W$  يمكن وضعها في شكل مصفوفة ارتباط . أي أن:

$$W'W = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1k} \\ r_{12} & 1 & r_{23} & \dots & r_{2k} \\ r_{13} & r_{23} & 1 & \dots & r_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{1k} & r_{2k} & r_{3k} & \dots & 1 \end{bmatrix}$$

حيث:

$$r_i = \frac{\sum_{u=1}^n (x_{iu} - \bar{x}_i)(x_{ju} - \bar{x}_j)}{(S_{ii} S_{jj})^{\frac{1}{2}}}$$

$$= \frac{S_{ij}}{(S_{ii} S_{jj})^{\frac{1}{2}}}, i, j = 1, 2, \dots, k$$

هو معامل الارتباط البسيط بين المتغير  $x_i$  ،  $x_j$  بنفس الشكل:

$$W'y^0 = \begin{bmatrix} r_{y1} \\ r_{y2} \\ r_{y3} \\ \vdots \\ r_{yk} \end{bmatrix}$$

حيث:

$$r_{yi} = \frac{\sum_{u=1}^n (x_{iu} - \bar{x}_i)(y_u - \bar{y})}{(S_{ii} S_{YY})^{\frac{1}{2}}} = \frac{s_{iy}}{(S_{ii} S_{YY})^{\frac{1}{2}}}$$

هو معامل الارتباط البسيط بين المتغير  $x_i$  ومتغير الاستجابة  $Y$ . عند استخدام الأسلوب الأول فإن  $Z'Z$  ترتبط بشدة بـ  $W'W$  حيث:

$$Z'Z = (n-1)W'W \quad (٤٨-٣)$$

في الحقيقة فإن التقديرات لمعاملات الانحدار في (٤٦-٣) و (٤٧-٣) متكافئة

ومتساوية أي أن الأسلوبين يعطيان نفس قيمة معاملات الانحدار  $b^*$ .

عادة يسمى متجه معاملات الانحدار  $b^*$  بمعاملات الانحدار المعيارية.

العلاقة بين المعاملات الأصلية والمعيارية كالتالي:

$$b_i = b_i^* \left( \frac{s_y}{s_i} \right), \quad i = 1, 2, \dots, k$$

و:

$$b_0 = \bar{y} - \sum_{i=1}^k b_i \bar{x}_i.$$

## مثال (٩-٤)

للمثال (١-٣) يمكننا استخدام معاملات الانحدار المعياري لمعرفة أي المتغيرين (الدخل وحجم الأسرة) أكثر تأثيراً على الإنجابية (الإستهلاك). بما أن قيم الانحراف المعياري للمتغيرات الإستهلاك والدخل وحجم الأسرة على التوالي هي:

$$s_y = \sqrt{\frac{1}{n-1} \left[ \sum y_i^2 - \frac{(\sum y_j)^2}{n} \right]}$$

$$= \sqrt{\frac{1}{9} \left[ 3396 - \frac{(180)^2}{10} \right]} = 4.163,$$

$$s_1 = \sqrt{\frac{1}{n-1} \left[ \sum x_{1j}^2 - \frac{(\sum x_{1j})^2}{n} \right]}$$

$$= \sqrt{\frac{1}{9} \left[ 406 - \frac{(60)^2}{10} \right]} = 2.261,$$

$$s_2 = \sqrt{\frac{1}{n-1} \left[ \sum x_{2j}^2 - \frac{(\sum x_{2j})^2}{n} \right]}$$

$$= \sqrt{\frac{1}{9} \left[ 182 - \frac{(40)^2}{10} \right]} = 1.563,$$

وعلى ذلك فإن معاملي الانحدار المعياري هما:

$$b_1^* = \frac{b_1 s_1}{s_y} = 2.363 \left( \frac{2.261}{4.163} \right) = 1.283,$$

$$b_2^* = \frac{b_2 s_2}{s_y} = -1.024 \left( \frac{1.563}{4.163} \right) = -0.3844.$$

أي نموذج الانحدار المقدر هو:

$$\hat{y}^* = 1.2883z_1^* - 0.3844z_2^*$$

أي أن زيادة الدخل بإنحراف معياري واحد تؤدي إلى زيادة في الإستهلاك بمقدار 1.283 بوحدات الإنحراف المعياري مع ثبات حجم الأسرة . كما أن زيادة حجم الأسرة بمقدار إنحراف معياري تؤدي إلى نقص الإستهلاك بمقدار 0.3844 بوحدات الإنحراف المعياري مع ثبات الدخل.

### (٣-١) معامل الارتباط الجزئي من الرتبة الأولى

يقاس معامل الارتباط الجزئي قوة وإتجاه العلاقة الخطية بين متغيرين بعد عزل أو إستبعاد أثر المتغيرات الأخرى. فمثلا  $r_{y1.2}$  يعنى الارتباط الجزئي بين  $y$  ,  $x_1$  بعد حذف تأثير  $x_2$  . ويسمى معامل الارتباط في هذه الحالة بمعامل الارتباط للجزئي من الرتبة الأولى حيث يساوى رتبة المعامل عدد المتغيرات المستبعد لثورها.

وبصورة عامة فإن معامل الارتباط الجزئي من الرتبة الأولى بين المتغيرين  $i$  ,  $j$  بعد جعل المتغير  $k$  ثابتاً هو:

$$r_{ij.k} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{(1-r_{ik}^2)(1-r_{jk}^2)}}$$

حيث  $r_{ij}$  هو معامل الارتباط البسيط بين المتغيرين  $i$  ,  $j$  و  $r_{ik}$  معامل الارتباط البسيط بين المتغيرين  $i$  ,  $k$  وهكذا .

فإذا كنا نرغب في إيجاد معامل الارتباط الجزئي بين المتغير  $Y$  والمتغير  $x_1$  مع إستبعاد أثر المتغير  $x_2$  فإن معامل الارتباط الجزئي ، يرمز له بالرمز  $r_{y1.2}$  هو:

$$r_{y1.2} = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{(1-r_{y2}^2)(1-r_{12}^2)}}$$

حيث  $r_{y1}$  هو معامل الارتباط البسيط بين المتغير  $Y$  والمتغير  $x_1$  وبالمثل يكون  $r_{y2}$ . أيضاً معامل الارتباط الجزئي بين المتغير  $Y$  والمتغير  $x_2$  مع إستبعاد أثر المتغير  $x_1$  ، يرمز له بالرمز  $r_{y2.1}$  ، هو:

$$r_{y2.1} = \frac{r_{y2} - r_{y1}r_{12}}{\sqrt{(1-r_{y1}^2)(1-r_{12}^2)}}$$

معامل الارتباط الجزئي يمكن أن يكون موجبا أو سالبا حيث تقع قيمته في الفترة  $[-1, 1]$  ويأخذ إشارة المعلمة المناظرة. لحساب معاملات الارتباط الجزئية من مثال (١-٣) نقوم أولا بحساب معاملات الارتباط البسيطة التالية:

$$\begin{aligned} r_{y1} &= \frac{\sum x_{1j}y_j - \frac{\sum x_{1j} \sum y_j}{n}}{\sqrt{\left[ \sum x_{1j}^2 - \frac{(\sum x_{1j})^2}{n} \right] \left[ \sum y_j^2 - \frac{(\sum y_j)^2}{n} \right]}} \\ &= \frac{1159 - \frac{(60)(180)}{10}}{\sqrt{\left[ 406 - \frac{(60)^2}{10} \right] \left[ 3396 - \frac{(180)^2}{10} \right]}} = 0.9325795, \\ r_{y2} &= \frac{\sum x_{2j}y_j - \frac{\sum x_{2j} \sum y_j}{n}}{\sqrt{\left[ \sum x_{2j}^2 - \frac{(\sum x_{2j})^2}{n} \right] \left[ \sum y_j^2 - \frac{(\sum y_j)^2}{n} \right]}} \\ &= \frac{766 - \frac{(40)(180)}{10}}{\sqrt{\left[ 182 - \frac{(40)^2}{10} \right] \left[ 3396 - \frac{(180)^2}{10} \right]}} = 0.785207, \end{aligned}$$

$$r_{12} = \frac{\sum x_{1j}x_{2j} - \frac{\sum x_{1j}\sum x_{2j}}{n}}{\sqrt{\left[\sum x_{1j}^2 - \frac{(\sum x_{1j})^2}{n}\right]\left[\sum x_{2j}^2 - \frac{(\sum x_{2j})^2}{n}\right]}}$$

$$= \frac{269 - \frac{(60)(40)}{10}}{\sqrt{\left[406 - \frac{(60)^2}{10}\right]\left[182 - \frac{(40)^2}{10}\right]}} = 0.9116072.$$

وعلى ذلك فإن :

$$r_{y1.2} = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{(1-r_{y2}^2)(1-r_{12}^2)}}$$

$$= \frac{0.9325795 - (0.785207)(0.9116072)}{\sqrt{1 - (0.7895207)^2} \sqrt{1 - (0.9116072)^2}} = 0.859283884.5,$$

$$r_{y2.1} = \frac{r_{y2} - r_{y1}r_{12}}{\sqrt{(1-r_{y1}^2)(1-r_{12}^2)}}$$

$$= \frac{0.785207 - (0.9325795)(0.9116072)}{\sqrt{1 - (0.9325795)^2} \sqrt{1 - (0.9116072)^2}} = -0.4376576.$$

يدعى مربع معامل الارتباط الجزئي بمعامل التحديد . يقيس معامل التحديد الجزئي المساهمة الهامشية لمتغير واحد من المتغيرات المستقلة ، عندما تكون جميع المتغيرات الأخرى موجودة أصلاً في النموذج . كثيراً ما تستخدم معاملات الارتباط الجزئي في التطبيقات العملية ، مع أنها لا تمتلك معنى واضح كوضوح معاملات التحديد الجزئية .

### (١٥-٣) معامل الارتباط الجزئي من الرتبة الثانية

يعتبر معامل الارتباط الجزئي من الرتبة الثانية إمتداداً لمعامل الارتباط الجزئي من الرتبة الأولى. فعلى سبيل المثال فإن معامل الارتباط الجزئي بين المتغيرين  $Y$  ,  $x_1$  بعد إستبعاد أثر المتغيرين  $x_2$  ,  $x_3$  يأخذ الصيغة التالية:



$$r_{y1,23} = \frac{r_{y1,3} - r_{12,3}r_{y2,3}}{\sqrt{(1-r_{12,3}^2)(1-r_{y2,3}^2)}}$$

وبصورة عامة معامل الارتباط الجزئي بين المتغيرين  $i, j$  بعد جعل تأثير بقية المتغيرات  $\ell, k$  ثابتة هو:

$$r_{ijk\ell} = \frac{r_{ijk} - r_{\ell k}r_{j\ell k}}{\sqrt{(1-r_{\ell k}^2)(1-r_{j\ell k}^2)}}$$

$$= \frac{r_{ij\ell} - r_{ik\ell}r_{jk\ell}}{\sqrt{(1-r_{ik\ell}^2)(1-r_{jk\ell}^2)}}$$

وبصورة عامة فإن معامل الارتباط الجزئي بين المتغيرين  $i, j$  بعد جعل جميع تأثيرات المتغيرات الأخرى ثابتة هو:

$$r_{ij} = \frac{-c_{ij}}{\sqrt{c_{ii} c_{jj}}} \quad (\text{كل المتغيرات الأخرى})$$

حيث  $c_{ij}$  ,  $c_{jj}$  ,  $c_{ii}$  هي عناصر المصفوفة العكسية لمصفوفة الارتباط. فعلى سبيل المثال إذا كان عدد المتغيرات المستقلة  $k = 3$  فعادة نحصل على مصفوفة الارتباط  $R$  لجميع المتغيرات المستقلة هي :

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} & r_{y1} \\ r_{12} & 1 & r_{23} & r_{y2} \\ r_{13} & r_{32} & 1 & r_{y3} \\ r_{y1} & r_{y2} & r_{y3} & 1 \end{bmatrix}$$

ثم نحصل على  $R^{-1}$  وهي :

$$R^{-1} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{y1} \\ c_{12} & c_{22} & c_{23} & c_{y2} \\ c_{13} & c_{32} & c_{33} & c_{y3} \\ c_{y1} & c_{y2} & c_{y3} & c_{yy} \end{bmatrix}$$

فعلى سبيل المثال:

-۳۱۲-

$$r_{23.y1} = \frac{-c_{23}}{\sqrt{c_{22} c_{33}}},$$

$$r_{y1.23} = \frac{-c_{y1}}{\sqrt{c_{11} c_{yy}}}.$$

**الفصل الرابع**  
**المخالفات في فروض نموذج الإتحدار الخطي المتعدد :**  
**كيفية اكتشافها وتصحيحها**  
**Violation in Multiple Regression Models: It's**  
**Detection and Correction**

مقدمه	(١-٤)
رسوم البواقي	(٢-٤)
رسوم البواقي الجزئية	(٣-٤)
رسوم الإتحدار الجزئي	(٤-٤)
البواقي المعيارية وبواقي ستوننت	(٥-٤)
استخدام مصفوفة القبة $H$ للتصرف على مشاهدات قاصية خاصة بالمتغيرات المستقلة	(٦-٤)
استخدام بواقي ستوننت المحذوفة للتعرف على مشاهدات قاصية خاصة بالمتغير التابع $Y$	(٧-٤)
تحديد المشاهدات المؤثرة	(٨-٤)
التأثير على القيم المقدرة	(١-٨-٤)
التأثير على معاملات الإتحدار	(٢-٨-٤)
مشكلة عدم الخطية ومعالجتها	(٩-٤)
مشكلة عدم تجانس الخطأ ومعالجتها	(١٠-٤)

#### (١-٤) مقلده

نكرنا في الفصل الثاني ، وعند تناولنا لنموذج الإنحدار البسيط أنه من المستحسن الكشف عن مخالفات فروض النموذج وذلك باستخدام تحليل البواقي أو اختبارات إحصائية معينة. أيضاً ناقشنا الطرق العلاجية لتصحيح هذه المخالفات. نفس الشيء يمكن تطبيقه في حالة الإنحدار الخطي المتعدد مع إجراء تعديلات صغيرة.

إن طرق الكشف عن المخالفات لفروض نموذج الإنحدار الخطي المتعدد تشمل الكشف عن المخالفات التالية :

١. دالة الإنحدار ليست خطية .
٢. حدود الخطأ ليست مرتبطه.
٣. النموذج ملائم لجميع المشاهدات بإستثناء مشاهدة واحدة أو قليل من المشاهدات القاصية .
٤. حدود الخطأ ليست طبيعية .
٥. حدود الخطأ ليس لها تباين ثابت .
٦. متغير مستقل مهم واحد أو عدد من المتغيرات المستقلة المهمة قد حذفت من النموذج .

#### (٢-٤) رسوم البواقي

البواقي  $e$  من نموذج الإنحدار المتعدد تلعب دور مهم في الحكم على صلاحية النموذج كما هو الحال في نموذج الإنحدار الخطي البسيط. رسوم البواقي في حالة الإنحدار الخطي البسيط يمكن تطبيقها مباشرة في الإنحدار المتعدد. هذا وهناك عدة رسوم مهمة للبواقي في تحليل الإنحدار المتعدد وهي :

- ١- رسم البواقي على ورق الإحتمال الطبيعي والذي يفيد في الكشف عما إذا كانت حدود الخطأ تتوزع بصورة طبيعيه وفق التوزيع الطبيعي.
- ٢- رسم البواقي مقابل القيم المقدره للإمتجابه  $\hat{y}$  حيث  $j = 1, 2, \dots, n$  والذي يفيد في تقييم صلاحية دالة الإنحدار وثبات تباين حدود الخطأ بالإضافة إلى تقديم معلومات عن المشاهدات القاصية (الخوارج) .
- ٣- رسم البواقي في التتبع الزمني إن وجد والذي يمكن أن يقدم معلومات حول ارتباطات ممكنة بين حدود الخطأ .

٤- رسم البواقي مقابل كل متغير مستقل  $x_i$  حيث  $i = 1, 2, \dots, k$  والذي يمكن أن يقدم معلومات إضافية حول صلاحية نموذج الإحدار بالنسبة لذلك المتغير المستقل (مثلاً قد نحتاج إلى تمثيل منحني لتأثير ذلك المتغير وحول تغيرات ممكنة في مقدار تبليين الخطأ فيما يتعلق بذلك المتغير المستقل).

٥- رسم البواقي مقابل متغيرات مستقلة مهمة حذفت من النموذج لرؤية ما إذا كان لهذه المتغيرات المحذوفة تأثيرات مهمة على المتغير التابع لم نعرف عليها بعد من خلال نموذج الإحدار. إن شكل الانتشار عند رسم البواقي مقابل المتغير المحذوف قد يشير إلى أن نموذج الإحدار المتعدد لأبعد أن يحتوي على هذا المتغير.

٦- رسم البواقي مقابل حدود التفاعل التي لم يشملها النموذج. مثل  $x_1 x_2$  و  $x_1 x_3$  و  $x_2 x_3$  وذلك لرؤية ما إذا كنا نحتاج، في النموذج، لبعض حدود التفاعل هذه أو لها جميعاً.

٧- رسم المتغير المستقل  $x_i$  مقابل المتغير المستقل  $x_{i'}$  و ( $i' \neq i$ ) والذي يفيد في دراسة العلاقة بين المتغيرات المستقلة وتشتت البيانات. عندما يكون هناك ارتباط قوي بين  $x_i, x_{i'}$  على الرسم فإن هذا يعني عدم ضرورة وجود المتغيرين  $x_i, x_{i'}$  معاً في النموذج. عندما يوجد متغيرين مستقلين بينهما علاقة قوية فإننا نقول أن هناك مشكلة تعدد العلاقات الخطية multicollinearity في البيانات. هذه المشكلة تؤثر على تقديرات المربعات الصغرى وتجعلها ليست ذات فائدة. سوف نناقش هذه المشكلة ويتفصيل أكثر في الفصل التاسع. رسم  $x_i$  مقابل  $x_{i'}$  مفيد أيضاً في اكتشاف النقاط البعيدة عن بقية النقاط والتي تؤثر على خواص النموذج. وبالإضافة إلى الرسوم السابقة هناك رسوم أخرى للبواقي سوف نناقشها باختصار.

### (٣-٤) رسوم البواقي الجزئية

هذه الرسوم تساعد في تحديد العلاقة بين البواقي والمتغير المستقل  $x_i$ . يعرف الباقي الجزئي للمتغير المستقل  $x_i$  كالتالي:

$$e_{ij}^* = y_j - b_1 x_{1j} - \dots - b_{i-1} x_{i-1,j} - b_{i+1} x_{i+1,j} - \dots - b_k x_{kj}$$

$$= e_j + b_j x_{ij}, \quad j=1,2,\dots,n.$$

رسم  $e_j^*$  مقابل  $x_{ij}$  يسمى رسم البواقي الجزئية ، هذه الرسوم قدمت من قبل  
Ezekiel and Fox (1959) و Larsen and McCleary (1972). وكما أن  
رسم البواقي  $e_j$  مقابل  $x_{ij}$  مفيد ، فإن رسم البواقي الجزئية مفيد في اكتشاف  
الملاحظات القاصية وعدم تجانس التباين. كثير من برامج الحاسب الآلي الجاهزة  
تنتج رسوم البواقي الجزئية.

#### مثال (١-٤)

في دراسة عن العلاقة بين امتصاص الماء في دقيق القمح و الخصائص  
المختلفة للدقيق وتحت فرض نموذج الحدار خطي متعدد تم الحصول على  
البيانات في جدول (١-٤) حيث  $Y$  تمثل كمية امتصاص الماء و( $x_1$ %)  
كمية البروتين و ( $x_2$ %) كمية النشا الذي يتعرض للفقء (التحطم مقياس  
بوحدة Farrand) والمطلوب إيجاد معادلة الإتحاد المقدرة ، ورسم :  
(أ) رسم  $x_2$  مقابل  $x_1$  (ب) رسم البواقي مقابل القيم المقدرة للاستجابة.  
(ج) رسم البواقي مقابل  $x_1$  و  $x_2$ . (د) حساب البواقي الجزئية.

جدول (١-٤)

$x_1$	$x_2$	$y$
8 . 5	2	30 . 9
8 . 9	3	32 . 7
10 . 6	3	36 . 7
10 . 2	20	41 . 9
9 . 8	22	40 . 9
10 . 8	20	42 . 9
11 . 6	31	46 . 3
12	32	47 . 2
12 . 5	31	44
10 . 4	28	47 . 7
1 . 2	36	43 . 9
11 . 9	28	46 . 8
11 . 3	30	46 . 2
13	27	47
12 . 9	24	46 . 8
12	25	45 . 9
12 . 9	28	48 . 8
13 . 1	28	46 . 2
11 . 4	32	47 . 8
13 . 2	28	49 . 2

## الحل

المصفوفتان  $X$  و  $y$  هما :

$$X = \begin{bmatrix} 1 & 8.5 & 2 \\ 1 & 8.9 & 3 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 13.2 & 28 \end{bmatrix} \quad y = \begin{bmatrix} 30.9 \\ 32.7 \\ \vdots \\ \vdots \\ 49.2 \end{bmatrix}$$

المصفوفة  $X'X$  هي :

$$X'X = \begin{bmatrix} 1 & 1 & \dots & \dots & 1 \\ 8.5 & 8.9 & \dots & \dots & 13.2 \\ 2 & 3 & \dots & \dots & 28 \end{bmatrix} \cdot \begin{bmatrix} 1 & 8.5 & 2 \\ 1 & 8.9 & 3 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & 13.2 & 28 \end{bmatrix}$$

$$= \begin{bmatrix} 20 & 218.2 & 478 \\ 218.2 & 2515.88 & 5271.8 \\ 478 & 5271.8 & 13322 \end{bmatrix}$$

والمتجه  $X'y$  هو :

$$X'y = \begin{bmatrix} 1 & 1 & \dots & \dots & 1 \\ 8.5 & 8.9 & \dots & \dots & 13.2 \\ 2 & 3 & \dots & \dots & 28 \end{bmatrix} \cdot \begin{bmatrix} 30.9 \\ 32.7 \\ \vdots \\ \vdots \\ 49.2 \end{bmatrix}$$

$$= \begin{bmatrix} 879.8 \\ 9710.06 \\ 21894.8 \end{bmatrix}$$

قيم  $b$  تعطى من المعادلة التالية :

$$b = (X'X)^{-1} X'y$$

حيث:

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 20 & 218.2 & 478 \\ 218.2 & 2515.88 & 5271.8 \\ 478 & 5271.8 & 13322 \end{bmatrix}^{-1} \begin{bmatrix} 879.8 \\ 9710.06 \\ 21894.8 \end{bmatrix}$$

$$= \begin{bmatrix} 1.12878 & -0.0762961 & -0.0103092 \\ -0.0762961 & 0.00748409 & -0.000224073 \\ -0.0103092 & -0.000224073 & 0.000533635 \end{bmatrix} \begin{bmatrix} 879.8 \\ 9710.06 \\ 21894.8 \end{bmatrix}$$

$$= \begin{bmatrix} 26.5433 \\ 0.63964 \\ 0.438 \end{bmatrix}$$

إن معادلة الانحدار المقدرة هي :

$$\hat{y} = 26.5433 + 0.63964x_1 + 0.438x_2$$

البواقي  $e_j$  معطاة في جدول (٢-٤) حيث  $e_j$  تحسب من العلاقة التالية :

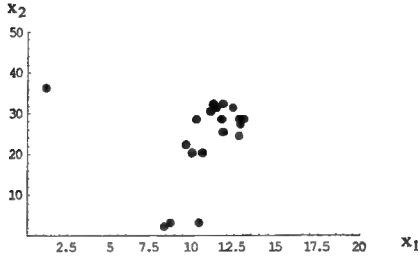
$$e_j = y_j - \hat{y}_j$$

جدول (٢-٤)

$y_j$	$\hat{y}_j$	$e_j$
30.9	32.8563	-1.95628
32.7	33.5501	-0.850131
36.7	34.6375	2.06248
41.9	41.8277	0.0723431
40.9	42.4478	-1.5478
42.9	42.2114	0.688559
46.3	47.5411	-1.24115
47.2	48.235	-1.035
44	48.1168	-4.11683
47.7	45.4596	2.24042
43.9	43.0789	0.821113
46.8	46.419	0.380958
46.2	46.9113	-0.711257
47	46.6846	0.315353
46.8	45.3067	1.49332
45.9	45.169	0.730992
48.8	47.0587	1.74132
46.2	47.1866	-0.986611
47.8	47.8512	-0.0512205
49.2	47.2506	1.94943

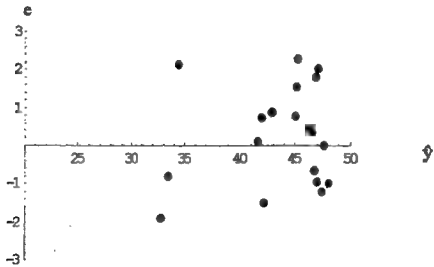


(أ) رسم  $x_2$  مقابل  $x_1$  موضح في شكل (١-٤) :



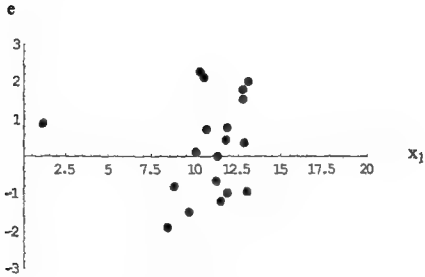
شكل (١-٤)

يتضح من شكل (١-٤) وجود بعض المشاهدات القاصية.  
(ب) رسم البواقي مقابل القيم المقدرة للاستجابة موضح في شكل (٢-٤) .



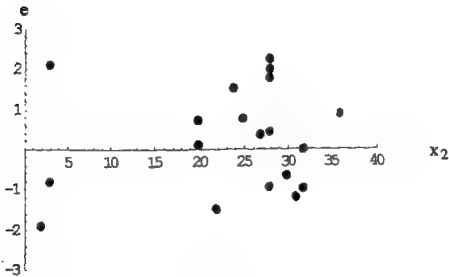
شكل (٢-٤)

(ج) رسم البؤالي مقابل  $x_1$  موضح في شكل (٣-٤) .



شكل (٣-٤)

رسم البؤالي مقابل  $x_2$  موضح في شكل (٤-٤) .



شكل (٤-٤)

(د) البيانات اللازمة لحساب البواقي الجزئية معطاة في جدول (٣-٤). كثير من برامج الحاسب الآلي الخاصة بالإنحدار تنتج رسوم البواقي الجزئية.

جدول (٣-٤)

$e$	$e_{jz}^*$	$e_{2j}^*$
-1.95628	3.48067	-1.08028
-0.850131	4.84267	0.463868
2.06248	8.84267	3.37648
0.0723431	6.59668	8.83234
-1.5478	4.72068	8.08819
0.688559	7.59668	9.44855
-1.24115	6.17868	12.3368
-1.035	6.64068	12.981
-4.11683	3.87868	9.46116
2.24042	8.89268	14.5044
0.821113	1.58868	16.5891
0.380958	7.99268	12.6449
-0.711257	6.51668	12.4287
0.315353	8.63068	12.1413
1.49332	9.74468	12.0053
0.730992	8.40668	11.681
1.74132	9.99268	14.0053
-0.986611	7.39268	11.2774
-0.0512205	7.24068	13.9648
1.94943	10.3927	14.2134

#### (٤-٤) رسوم الإنحدار الجزئي

للحصول على رسم الإنحدار الجزئي نحسب البواقي الناتجة من انحدار كلا من متغير الاستجابة  $Y$  والمتغير المستقل المعنى  $x_i$  على المتغيرات المستقلة الأخرى في نموذج الإنحدار . ورسم مجموعتي البواقي هاتين أحدهما في مقابل الأخرى يكشف عن طبيعة علاقة الإنحدار للمتغير المستقل  $x_i$  موضع الدراسة وكذلك الأهمية الهامشية لهذا المتغير في تخفيض تشتت البواقي. في صيغة مصقوفة سوف نكتب تلك الكميات (البواقي) على الشكل  $e_{x_i|x(i)}$  ،  $e_{y|x(i)}$  على التوالي حيث  $X_{(i)}$  هي المصفوفة الأصلية  $X$  مع حذف المتغير المستقل  $x_i$  . لإثبات كيف تعرف تلك الكميات، ليكن النموذج :

$$Y = X\beta + \varepsilon$$

$$= X_{(i)}\beta + x_i\beta_i + \varepsilon \quad (1-4)$$

وبضرب طرفي المعادلة (1-4) في  $(I - H_{(i)})$

نحصل على:

$$(I - H_{(i)})Y = (I - H_{(i)})X_{(i)}\beta + (I - H_{(i)})x_i\beta_i + (I - H_{(i)})\varepsilon$$

وبما أن  $(I - H_{(i)})X_{(i)} = 0$  فإن:

$$(I - H_{(i)})Y = (I - H_{(i)})x_i\beta_i + (I - H_{(i)})\varepsilon$$

أو :

$$e_{Y|X_{(i)}} = \beta_i e_{x_i|X_{(i)}} + \varepsilon^*$$

حيث  $\varepsilon^* = (I - H_{(i)})\varepsilon$  . وهذا يعني أن رسم الانحدار الجزئي يكون له ميل  $\beta_i$ .

وعلى ذلك إذا دخلت  $x_i$  الانحدار في شكل خطي فإن رسم الانحدار الجزئي يوضح علاقة خطية تمر بنقطة الأصل. كثير من برامج الحاسب الآلي الخاصة بالانحدار (مثل SAS) تحسب رسوم الانحدار الجزئي.

#### (٥-٤) البوابي المعيارية وبوابي ستيوننت

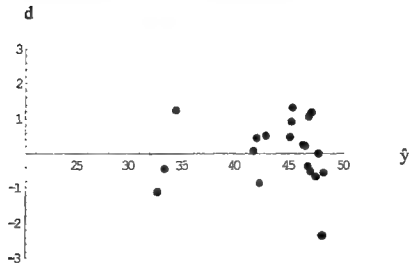
في الفصل الثاني تناولنا نوعين من البوابي هما البوابي المعيارية وبوابي ستيوننت وذلك للكشف عن مشاهدات قاصية في قسم  $Y$ . تعرف البوابي المعيارية كالآتي:

$$d_j = \frac{e_j}{\sqrt{MSE}} \quad , j = 1, 2, \dots, n.$$

للمثال (1-4) فإن رسم البوابي المعيارية  $d_j$  مقابل  $\hat{y}_j$  معطاة في شكل (٥-٤) وذلك من البيانات المعطاة في جدول (٤-٤).

جدول (٤-٤)

$\hat{y}_j$	$d_j$
32.8563	-1.16113
33.5501	-0.504588
34.6375	1.22417
41.8277	0.0429386
42.4478	-0.918683
42.2114	0.408688
47.5411	-0.736673
48.235	-0.614318
48.1168	-2.44351
45.4596	1.32978
43.0789	0.487365
46.419	0.226114
46.9113	-0.422161
46.6846	0.187175
45.3067	0.886345
45.169	0.433874
47.0587	1.03354
47.1866	-0.585594
47.8512	-0.0304015
47.2506	1.15706



شكل (٤-٥)

في حالة الإتحدار الخطي المتعدد، علمنا من البند (٣-٧) أننا يمكننا كتابة  
متجه البواقي  $e$  كالآتي :

$$e = (I - H)Y,$$

حيث  $H = X(X'X)^{-1}X'$  ترمز لمصفوفة القبة والتي تلعب دوراً مهماً في أي  
دراسة للبواقي وفي مواضيع متقدمة من الإتحدار. وكما قلنا سابقاً فإن  $H$  تولد  
القيم المقدرة عند ضربها في متجه القيم المشاهدة للاستجابة. أي أن :

$$\hat{y} = Hy,$$

حيث  $\hat{y}$  هو المتجه الذي عنصره رقم  $\hat{y}_j$  هو  $\hat{y}_j$ . العنصر  $h_{jj}$  على القطر  
الرئيسي للمصفوفة  $H$  يسمى قيمة القبة  $\hat{y}$  أو الرافعة (الوعزم) حيث  
يقيس المسافة بين المشاهدة رقم  $j$  ( $x_{1j}, x_{2j}, \dots, x_{kj}$ ) ومتوسط قيم كل الحالات  
( $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ ). وبالتعويض عن  $Y$  في المعادلة (٢-٤) بقيمتها  
وهي  $Y = XB + \varepsilon$  نحصل على :

$$\begin{aligned} e &= (I - H)(XB + \varepsilon) \\ &= XB - HXB + (I - H)\varepsilon \\ &= XB - X(X'X)^{-1}X'XB + (I - H)\varepsilon \\ &= (I - H)\varepsilon. \end{aligned}$$

أي إن البواقي تمثل التحويلة الخطية للملاحظات وأيضاً للأخطاء  $\varepsilon$ .  
مصفوفة التغاير للبواقي هي :

$$\begin{aligned} \text{Cov}(e) &= \text{Cov}[(I - H)\varepsilon] \\ &= (I - H)\text{Cov}(\varepsilon)(I - H)' \\ &= \sigma^2(I - H). \end{aligned}$$

ونذلك لأن  $\text{Cov}(\varepsilon) = \sigma^2 I$  و  $(I - H)$  متماثلة وجامعة. عموماً  
المصفوفة  $(I - H)$  ليست قطرية وعلى ذلك البواقي مرتبطة ولها تباينات مختلفة  
التباين للبواقي رقم  $j$  هو :

$$\text{Var}(e_j) = \sigma^2(1 - h_{jj}),$$

حيث  $h_{jj}$  هو العنصر على القطر الرئيسي للمصفوفة  $H$  و  $0 \leq h_{jj} \leq 1$ .  
ولقد افترض كوكوين باعاً الاختلاف في تباين البواقي في حساب البواقي ومن ثم افترضت بواقي  
ستودنت و المعرفة كالآتي :

$$r_j = \frac{e_j}{\sqrt{\text{MSE}(1 - h_{jj})}}, \quad j = 1, 2, \dots, n.$$

وتباين البواقي  $r_j$  ثابت حيث  $\text{Var}(r_j) = 1$ . للفتات الكبيرة من البيانات يوجد فروق بسيطة بين البواقي المعيارية وبواقي مستويكنت. التفسير بين  $e_j, e_{j'}$  حيث  $j \neq j'$  هو :

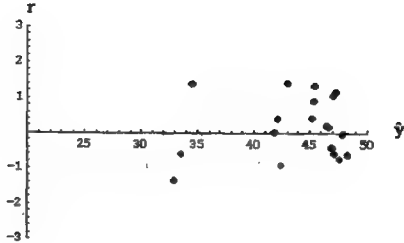
$$\text{Cov}(e_j, e_{j'}) = -\sigma^2 h_{jj}.$$

للمثال (٦-٤) قيم كل من  $h_{jj}$  والبواقي  $r_j$  معطاة في جدول (٦-٤).

جدول (٦-٤)

$e_j$	$h_{jj}$	$r_j$
-1.95628	0.325752	-1.41407
-0.850131	0.294507	-0.600746
2.06248	0.280913	1.44361
0.0723431	0.0606484	0.0443031
-1.5478	0.0602024	-0.947652
0.688559	0.0580149	0.421085
-1.24115	0.0782682	-0.767313
-1.035	0.0899469	-0.643962
-4.11683	0.0907619	-2.56256
2.24042	0.0618541	1.37292
0.821113	0.886413	1.44607
0.380958	0.0644865	0.233777
-0.711257	0.0699287	-0.437743
0.315353	0.0849159	0.195667
1.49332	0.0795539	0.923854
0.730992	0.0590002	0.447269
1.74132	0.0849517	1.08046
-0.986611	0.0908409	-0.614154
-0.0512205	0.08503	-0.0317828
1.94943	0.0940101	1.21561

رسم البواقي  $r_j$  مقابل  $\hat{y}$  معطاة في شكل (٦-٤).



شكل (٦-٤)

(٦-٤) استخدام مصفوفة القيمة  $H$  للتعرف علي مشاهدات قاصية خاصة بالمتغيرات المستقلة

يعتبر العنصر القطري  $h_{jj}$  في مصفوفة القيمة  $H$  مؤشر مفيد للكشف عن المشاهدات القاصية الخاصة بالمتغيرات المستقلة. وذلك في دراسة متعددة المتغيرات، حيث يقيس المسافة بين الملاحظة رقم  $j$  ( $x_{1j}, x_{2j}, \dots, x_{kj}$ ) ومتوسط قيم كل الحالات ( $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ ). وهكذا يشير كبر قيمة العزم  $h_{jj}$  إلى أن الملاحظة  $j$  بعيدة عن مركز المشاهدات جميعا. وعادة تعتبر قيمة  $h_{jj}$  كبيرة إذا تجاوزت ضعف متوسط قيم  $h_{jj}$  [Belsley et al (1980)] ونرمز له بالرمز  $h$  حيث:

$$h = \frac{\sum h_{jj}}{n} = \frac{p}{n}$$

وذلك لأن مجموع قيم العناصر القطرية للمصفوفة  $H$  يساوي عدد معالم نموذج الانحدار الخطي بما في ذلك المعامل الثابت. وبالتالي فإن قيم  $h_{jj}$  التي تزيد عن



$2p/n$  تعتبر وفقا لهذه القاعدة مؤشرا لوجود مشاهدات قاصية خاصة بالمتغيرات المستقلة. ايضا اقترح [Neter et al (1990)] أن الرافعه التي تزيد قيمتها عن 0.5 كبيره وتشير إلى أن الحالة شاذه وتستدعي دراستها. ولأنه في العينات الصغيرة يتوقع ترشيح عدد كبير من الحالات الشاذة لفحصها فهناك اقتراح بدراسة كل الحالات التي تزيد قيم رافعتها ( $h_{jj}$ ) عن ثلاثة اضعاف متوسط قيمة الرافعات  $\left(h_{jj} > \frac{3(p)}{n}\right)$  بدلا من دراسة الحالات التي تزيد قيم رافعتها  $h_{jj}$  عن ضعف متوسط قيم الرافعات  $\left(h_{jj} > \frac{2(p)}{n}\right)$ .

للمثال (٤-١) يعطى جدول (٤-٦) قيم الرافعة  $h_{jj}$  ويتضح من الجدول أن هناك حالتين تزيد قيم رافعتها عن ضعف متوسط قيم الرافعات  $\left(\frac{2p}{20} = \frac{2(3)}{20} = 0.3\right)$  وهي (1) و (11). هذا وقد بلغت قيم الرافعات المناظرة لهذه الحالات على التوالي 0.325752 , 0.886413. ايضا نجد أن الحالة رقم (11) هي الحالة الوحيدة التي تزيد قيمة رافعتها عن ثلاثة اضعاف متوسط قيم الرافعات (0.45). كما يلاحظ وجود حالة قاصيه ولحده حسب اقتراح [Neter et al (1990)] وهي الحالة رقم (11) والتي تزيد عن 0.5.

(٤-٧) استخدام بواقي ستوبونت المحذوفه للتعرف على مشاهدات قاصيه خاصه بالمتغير التابع Y

تستخدم البواقي المحذوفه للكشف عن مشاهدات المتغير التابع القاصيه . ويعرف الباقي المحذوف للمشاهده  $j$  بأنه الفرق بين قيمة  $y_j$  الفعلية والقيمة المقدرة لها  $y(j)$  باستخدام نموذج الإنحدار الذي يتم تقديره بعد استبعاد المشاهده رقم  $j$  . وهذه الطريقة تكرر لكل مشاهده  $j$  حيث  $j = 1, 2, \dots, n$ ، والتي تنتج فئة من  $n$  من البواقي المحذوفه  $e_{(1)}, e_{(2)}, \dots, e_{(n)}$  . ومن هذا يتضح أننا نحتاج إلى بناء عدد  $n$  من نماذج الإنحدار المقدرة لحساب البواقي المحذوفه لكل المشاهدات . ومن حسن الحظ فإنه يمكن حساب البواقي المحذوفه من معادله جبرية دون الحاجة لبناء هذا العدد من النماذج (1990) Neter et al حيث:

$$e_{(j)} = \frac{e_j}{1 - h_{jj}}$$

تباين البواقي المحذوفة سوف يكون:

$$\begin{aligned}\text{Var}(e_{(j)}) &= \text{Var}\left[\frac{e_j}{1-h_{jj}}\right] \\ &= \frac{1}{(1-h_{jj})^2} \sigma^2 (1-h_{jj}) \\ &= \frac{\sigma^2}{1-h_{jj}}.\end{aligned}$$

وعلى ذلك البواقي المحذوفة المعيارية سوف تكون :

$$\begin{aligned}\frac{e_{(j)}}{\sqrt{\text{Var}(e_{(j)})}} &= \frac{e_j / (1-h_{jj})}{\sqrt{\sigma^2 / (1-h_{jj})}} \\ &= \frac{e_j}{\sqrt{\sigma^2 (1-h_{jj})}}.\end{aligned}\quad (٣-٤)$$

والتي بعد استبدال  $\sigma^2$  بـ MSE نحصل على بواقي ستوننت التي ناقشناها سابقا .

بإبقاء ستوننت  $r_j$  الذي ناقشناه في البند (٥-٤) يعتبر أداة لاكتشاف المشاهدات القاصية . عادة يستخدم MSE لتقدير  $\sigma^2$  في حساب  $r_j$  . هناك أسلوب آخر يمكن استخدامه في تقدير  $\sigma^2$  والذي يعتمد على  $s_{(j)}^2$  حيث :

$$s_{(j)}^2 = \frac{(n-p)\text{MSE} - e_j^2 (1-h_{jj})}{n-p-1} \quad (٤-٤)$$

والذي يستخدم لتقدير  $\sigma^2$  في (٣-٤) بدلا من MSE وذلك للحصول على باقي يسمى باقي ستوننت المحذوف حيث :

$$t_j = \frac{e_j}{\sqrt{s_{(j)}^2 (1-h_{jj})}}, \quad j=1,2,\dots,n.$$

تحت الفروض القياسية فإن  $t_j$  يتبع توزيع  $t$  بدرجات حرية  $n-p-1$  . تعتبر بواقي ستوننت المحذوفة طريقة مناسبة لاكتشاف المشاهدات القاصية الخاصة

بالمتغير التابع . بواقي ستوننت المحذوفه معطاة في جدول (٧-٤) والخاصه بالمثال (١-٤).

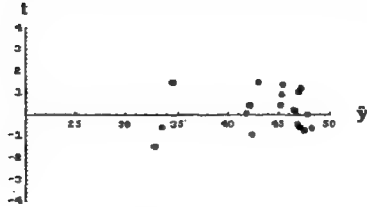
وإذا اعتبرنا الذيلين على الجانبين لتوزيع  $t$  بمساحه 0.05 لكل منهما متطرفين ، فنحتاج لمقارنه قيم بواقي ستوننت بقيمه توزيع  $t$  بدرجات حريه  $n-p-1$  ، أي المقارنه بـ  $t_{0.05}(16) = 1.746$  على وجه التحديد

وبما أن المشاهده  $y = 44$  لها باقي ستوننت المحذوف  $t_0 = -3.1734$  وبما أن  $|t_0| = 3.1734 > 1.746$  فإننا نعتبر أن المشاهده  $y$  مشاهده قاصيه.

جدول (٧-٤)

$e_j$	$h_{jj}$	$t_j$
-1.95628	0.325752	-1.46043
-0.850131	0.294507	-0.589096
2.06248	0.280913	1.49515
0.0723431	0.0606484	0.0429828
-1.5478	0.0602024	-0.944647
0.688559	0.0580149	0.41066
-1.24115	0.0782682	-0.757639
-1.035	0.0899469	-0.632497
-4.11683	0.0907619	-3.1734
2.24042	0.0618541	1.41254
0.821113	0.886413	1.49805
0.380958	0.0644865	0.227163
-0.711257	0.0699287	-0.427087
0.315353	0.0849159	0.190039
1.49332	0.0795539	0.919654
0.730992	0.0590002	0.436491
1.74132	0.0849517	1.08615
-0.986611	0.0908409	-0.602538
-0.0512205	0.08503	-0.0308347
1.94943	0.0940101	1.23418

رسم بواقي ستويونت المحذوفة مقابل  $\hat{y}_j$  معطاة في شكل (٧-٤).



شكل (٧-٤)

#### (٨-٤) تحديد المشاهدات المؤثرة

بعد تحديد المشاهدات القاصية بالنسبة لقيمها في المتغيرات المستقلة و (أو) قيمتها في متغير الاستجابة تكون الخطوة التالية هو التعرف على ما إذا كانت هذه المشاهدات القاصية مؤثرة (influential) أم لا؟

وتعتبر الملاحظة مؤثرة إذا كان استبعادها يحدث تغيراً ملحوظاً في قيم نموذج الانحدار والإحصاءات المرتبطة بها. وسوف نناقش هنا مقاييس للتأثير وهي مقاييس مستخدمة على نطاق واسع في التطبيق العملي ويعتمد كل مقياس على حذف مشاهدة واحدة لقياس تأثيرها.

#### (١-٨-٤) التأثير على القيم المقدرة

لقياس تأثير الملاحظة  $j$  على القيمة المقدرة سوف نستخدم المقياس التالي:

$$DFFITS_j = \frac{\hat{y}_j - \hat{y}_{(j)}}{\sqrt{s_{(j)}^2 h_{jj}}}, \quad j = 1, 2, \dots, n.$$

ويرمز الحرفان DF للفرق بين القيمة المقدرة  $\hat{y}_j$  للمشاهدة  $y_j$  عند استخدام جميع المشاهدات  $n$  في إيجاد معادلة الانحدار المقدرة وبين القيمة المقدرة  $\hat{y}_{(j)}$  التي نحصل عليها عند حذف المشاهد  $j$  في عملية تقدير معادلة الانحدار المقدرة. المقام في صيغة  $DFFITS_j$  يمثل معياره وذلك لأن  $Var(\hat{y}_j) = \sigma^2 h_{jj}$ . وعلى ذلك  $DFFITS_j$  يمثل عدد الانحرافات المعيارية والتي تتغيرها القيمة المقدرة  $\hat{y}_j$  عند حذف المشاهد  $j$ . ويمكن حساب  $DFFITS_j$  مستخدمين فقط النتائج المتوفرة من تقدير مجموع البيانات بكاملها وذلك تبعا للعلاقة التالية:

$$DFFITS_j = \left( \frac{h_{jj}}{1 - h_{jj}} \right)^{\frac{1}{2}} \frac{e_j}{[s^2_{(j)}(1 - h_{jj})]^{\frac{1}{2}}}$$

$$= \left( \frac{h_{jj}}{1 - h_{jj}} \right)^{\frac{1}{2}} t_j$$

حيث  $t_j$  تمثل باقي ستويونت المحذوف. وعلى ذلك  $DFFITS_j$  يمثل قيمة من قيم باقي ستويونت المحذوف بعد ضربها في  $[h_{jj}/(1 - h_{jj})]^{\frac{1}{2}}$ . وعلى ذلك فإن زياده أو نقصان قيمة  $DFFITS_j$  يتم من خلال عامل ما في واقع الأمر داله في قيم  $h_{jj}$ . وإذا كانت المشاهده  $j$  قاصيه في قيم المتغيرات المستقلة ولها قيمة  $h_{jj}$  مرتفع فإن هذا العامل سيكون أكبر من الواحد الصحيح وبالتالي يؤدي إلى أن القيمة المطلقة لـ  $DFFITS_j$  كبيره. وحسب Belsley et al (1980) تعتبر الحالة مؤثره إذا كانت قيمة  $DFFITS_j$  المطلقة أكبر من  $2\sqrt{\frac{p}{n}}$ .

وحسب Netel et al (1990) تعتبر الحالة مؤثره إذا كانت القيمة المطلقة لـ  $DFFITS_j$  أكبر من واحد صحيح في حالة العينات الصغيره والمتوسطه وإذا كانت أكبر من  $2\sqrt{\frac{p}{n}}$  في حالة العينات الكبيره. واخيرا حسب Chatterjee and Hadi (1988) تعتبر الحالة مؤثره إذا كانت القيمة المطلقة لـ  $DFFITS_j$  أكبر قليلا من التي اقترحتها Belsley et al (1980)، أي أن:  $|DFFITS_j| > 2\sqrt{\frac{p}{n-p}}$ .

(٢-٨-٤) التأثير على معاملات الإحدار

مقياس الأثر على كل معاملات الإحدار (مقياس كوك)

لقد اقترح [Cook (1979)] مقياس (مسافة كوك) لمربع المسافة بين تقدير المربعات الصغرى (b) المبني على كل المشاهدات التي عددها n والتقدير  $b_{(j)}$  الذي نحصل عليه بعد حذف المشاهدة رقم j. أي أنه مقياس إجمالي للتأثير المشترك للمشاهدة j على جميع معاملات الإحدار المقدرة. وهذا المقياس يعتمد على مفهوم منطقة الثقة المشتركة لمعاملات الإحدار  $b_j$  حيث  $i = 0, 1, \dots, k$ . وعلى ذلك فإن حدود منطقة الثقة المشتركة معطاه بالعلاقة التالية:

$$D_j(M, c) = \frac{(b_{(j)} - b)'M(b_{(j)} - b)}{p} = F_{\alpha}(p, n - p), j = 1, 2, \dots, n. \quad (٥-٤)$$

حيث  $c = pMSE$  و  $M = X'X$  وعلى ذلك مقياس مسافته كوك يصبح:

$$D_j(M, c) = \frac{(b_{(j)} - b)'X'X(b_{(j)} - b)}{p \text{ MSE}}, j = 1, 2, \dots, n.$$

ويمكن حساب مقياس مسافة كوك  $D_j$  بدون تقدير نموذج إحدار جديد في كل مرة تحذف فيها مشاهدة مختلفة والصيغة المكافئة جبرياً هي:

$$D_j = \frac{e_j^2}{p \text{ MSE}} \left[ \frac{h_{jj}}{(1 - h_{jj})^2} \right] \quad (٦-٤)$$

ونلاحظ من (٦-٤) أن  $D_j$  يعتمد على حجم الباقي  $e_j$  وقيمة الرافعة  $h_{jj}$ . وكلما كان أي من  $e_j$  أو  $h_{jj}$  أكبر كلما كان  $D_j$  أكبر. وهذا ويمكن أن تكون المشاهدات مؤثرة:

• إذا كان لدينا باقي  $e_i$  كبير وقيمة الرافعة معتدلة  $h_{ii}$

• إذا كان لدينا قيمة رافعة  $h_{jj}$  كبيرة مع باقي  $e_j$  من حجم معتدل

• إذا كان لدينا باقي  $e_j$  كبير وقيمة رافعة  $h_{jj}$  كبيرة.

ولتحديد أثر الملاحظة رقم  $j$  على معاملات الانحدار فإن هناك اقتراح بأن تتم مقارنة قيمة  $D_j$  بالقيمة  $4/(n-p)$  فإذا كانت قيمة  $D_j$  أكبر من هذه القيمة تعتبر الملاحظة  $j$  حالة مؤثرة على قيم معاملات نموذج الانحدار وإلا تعتبر الحالة غير مؤثرة. وبينما لا يتبع  $D_j$  توزيع  $F$  فقد وجد أنه من المفيد نسبة القيمة  $D_j$  إلى التوزيع  $F$  المقابل وفقا للمعادلة (٤-٥) ومعرفة المنين المرافق لتلك القيمة. وإذا كانت قيمة المنين أقل من 10 أو 20 بالمائة فيكون للملاحظة  $j$ ، على ما يبدو، تأثير بسيط على معاملات الانحدار. وعلى الوجه الآخر، إذا كانت قيمة المنين قرب الـ 50 بالمائة أو أكثر فينبغي اعتبار المسافة بين المتجهين  $b_{(j)}$ ،  $b$  كبيره مما يتضمن أن للملاحظة تأثيرا كبيرا على نموذج الانحدار.

#### مقياس الأثر على معاملات الانحدار (مقياس $DFBETAS_{ij}$ )

لقد اقترح Belsley et al (1980) إحصاء لقياس الفرق بين قيم معاملات الانحدار المقدرة باستخدام كل المشاهدات التي عددها  $n$  وقيم معاملات الانحدار المقدرة بعد حذف الملاحظة رقم  $(j)$  أي باستخدام  $(n-1)$  من المشاهدات. هذا الإحصاء يأخذ الشكل التالي:

$$DFBETAS_{ij} = \frac{b_i - b_{i(j)}}{\sqrt{s_{(j)}^2 c_{ii}}} \quad \text{و } i = 0, 1, \dots, p.$$

حيث  $c_{ii}$  هو العنصر القطري رقم  $i$  للمصفوفة  $(X'X)^{-1}$  و  $b_i$  هو معامل الانحدار رقم  $i$  المحسوب باستخدام كل الحالات  $(n)$ ،  $b_{i(j)}$  هو معامل الانحدار رقم  $i$  المحسوب بدون استخدام الملاحظة رقم  $j$ . القيم الكبيرة من  $DFBETAS_{ij}$  توضح أن الملاحظة رقم  $j$  يمكن اعتبارها مؤثرة على معامل الانحدار رقم  $i$ . وكمعيار عام لتحديد الحالات المؤثرة فقد اقترح Neter et al (1990) أنه إذا كانت قيمة  $|DFBETAS_{ij}|$  أكبر من الواحد الصحيح في حالات العينات الصغيرة أو أكبر من  $2/\sqrt{n}$  في حالة العينات الكبيرة تعتبر الحالة مؤثرة أي تعتبر الحالة رقم  $j$  مؤثرة إذا تحقق الشرط التالي:

في حالة العينات الصغيرة:

$$|DFBETAS_{i,j}| > 1$$

في حالة العينات الكبيرة:

$$|DFBETAS_{i,j}| > 2/\sqrt{n}.$$

وبلاحظ أنه لحساب  $DFBETAS_{i,j}$  لكل المشاهدات نحتاج لتقدير (n) نموذج انحدار.

البيانات في جدول (٨-٤) تمثل عينة عشوائية ذات حجم (n = 14) مشاهدات لكل من المتغير التابع Y والمتغيرات المستقلة  $x_1, x_2, x_3$ .

جدول (٨-٤)

y	$x_1$	$x_2$	$x_3$
3.33	0.276	0.240	0.625
3.51	0.249	0.254	0.512
3.55	0.249	0.249	0.488
3.65	0.260	0.245	0.524
3.80	0.271	0.250	0.588
4.20	0.241	0.252	0.475
4.22	0.269	0.254	0.513
4.27	0.264	0.270	0.463
4.31	0.270	0.274	0.512
4.48	0.240	0.264	0.405
4.53	0.259	0.280	0.450
4.55	0.252	0.266	0.480
4.62	0.258	0.268	0.456
5.86	0.293	0.286	0.506



أوجد قيم  $DFFITS_j$  و  $DFBETAS_j$  ومساافة كوك للمشاهدات وماذا تستنتج من تلك القيم؟

**الحل**

يعطى جدول (٩-٤) قائمة بقيم  $h_{jj}$ ,  $D_j$ ,  $DFFITS_j$ . بالنسبة لـ  $D_j$  المئين العشرين لتوزيع F بدرجات حرية (4, 10) (مأخوذ من الحزم الجاهزة الخاصة بالإحصاء لبرنامج Mathematica) يساوي 0.406574. يلاحظ أن كل قيم مسافه كوك أقل من هذه القيمة. وعلى ذلك لا يوجد أي حالة مؤثرة. في الحقيقة فإن المئين الخمسين هو 0.898817 وعلى ذلك لا توجد أي قيمة لـ  $D_j$  تقترب من المستوى الضروري لوجود حالة مؤثرة على نموذج الانحدار.

**جدول (٩-٤)**

$h_{jj}$	$D_j$	$DFFITS_j$
0.446993	0.0879027	0.575188
0.231595	0.0487779	0.433307
0.169291	0.0388601	-0.389163
0.182651	0.0621402	-0.501698
0.192002	0.0823222	0.586544
0.3769	0.222289	0.96858
0.174635	0.0893455	-0.622104
0.179229	0.108167	-0.696866
0.315062	0.0471378	0.420653
0.318829	0.0801743	-0.556649
0.371122	0.00283181	0.101065
0.153562	0.0923556	0.646137
0.119834	0.00632049	-0.152264
0.768295	0.0598614	0.465906

الآن بالنسبة لقيم  $DFFITs_j$  لا يوجد أي قيم مؤثرة وذلك لعدم وجود أي قيم لـ  $|DFFITs_j|$  تزيد عن 1. القيمة القريبة من 1 في جدول (٤-٩) هي 0.96858 والمقابلة للملاحظة العاشرة.

يعطى جدول (٤-١٠) قائمة بقيم الـ  $DFBETAs_{i,j}$  حيث  $i = 0, 1, 2, 3$  ,  $j = 1, 2, \dots, 14$  إلى 14.

يلاحظ من جدول (٤-١٠) عدم وجود قيم مطلقة لـ  $DFBETAs_{i,j}$  تزيد عن الواحد الصحيح. وعلى ذلك لا يوجد أي حالات لها تأثير معنوي على نموذج الاتحاد.

جدول (٤-١٠)

$b_0$	$b_1$	$b_2$	$b_3$
-0.0143844	-0.115759	0.382709	-0.136519
-0.0559135	-0.289016	-0.101049	0.209172
-0.123586	0.108709	0.126155	-0.00171137
-0.235483	-0.0653034	-0.0718679	0.257519
-0.0000106592	-0.0497338	0.358846	-0.156987
0.835436	0.62326	-0.674636	-0.59404
-0.226309	-0.307155	-0.220736	0.409338
0.486631	0.446869	-0.124454	-0.530767
-0.355183	-0.301588	0.150272	0.345162
-0.309542	-0.233555	0.470109	0.105511
-0.0680988	-0.059592	-0.00941033	0.0865301
0.25121	0.2822	-0.396119	-0.105235
-0.0263566	-0.0593643	0.0557525	0.0124995
-0.0907156	0.234257	0.199504	-0.10239

### (١٤-١) مشكلة عدم الخطية ومعالجتها

في البند (٢-٣) ناقشنا اختبار نقص التوفيق في حالة الانحدار الخطي البسيط. وقد كانت تشتمل الطريقة على تجزئة مربعات الخطأ (البواقي) إلى جزئين، جزء يعود إلى الخطأ الصافي والجزء الآخر يعود إلى نقص جودة التوفيق أي أن :

$$SSE = SSPE + SSLF,$$

حيث مجموع المربعات الصافي يحسب باستخدام استجابات عند مشاهدات مكررة عند نفس المستوى من  $x$ . هذه الطريقة تعطي تقدير لـ  $\sigma^2$  لا يعتمد على التوزيع.

الطريقة السابقة يمكن تعميمها إلى الانحدار الخطي المتعدد. حساب SSPE يتطلب تكرار مشاهدات  $y$  عند نفس الفئة من مستويات المتغيرات المستقلة  $x_1, x_2, \dots, x_k$ . أي أن بعض الصفوف من المصفوفة  $X$  سوف تكون نفسها. في الجزء التالي سوف نوضح خطوات تنفيذ هذه الطريقة بمثال.

### مثال (٤-٣)

أجريت تجربة لدراسة تأثير كل من درجة الحرارة  $x_1$  و مدة التخزين  $x_2$  على حامض الاسكريك ascorbic acid في الفاصوليا الخضراء والبيانات معطاة من جدول (٤-١١)

جدول (٤-١١)

$x_1 \backslash x_2$	المجموع	المشاهدات $y_{ij}$	المجموع $y_i$
-20	2	1	15,16,14 (45)
	4	2	17,15,15 (47)
	6	3	15,16,14 (45)
-15	2	1	15,15,16 (46)
	4	2	12,15,15 (42)
	6	3	13,15,14 (42)
-10	2	1	11,11,12 (34)
	4	2	11,9,8 (28)
	6	3	8,7,6 (21)
المجموع			350

من جدول (١١-٤) فإن :

مجموع المربعات الكلية SYY هو :

$$\begin{aligned} SYY &= \sum \sum y_{ij}^2 - \frac{(\sum y_{ij})^2}{n} \\ &= (15^2 + 16^2 + 14^2 + 17^2 + \dots + 6^2) - \frac{(350)^2}{27} \\ &= 4784 - 4537.0370 = 246.963. \end{aligned}$$

مجموع المربعات الذي يعود للمجاميع وسوف نرمز له بالرمز SSBT وهو :

$$\begin{aligned} SSBT &= \sum \left( \frac{\sum y_{ij}^2}{n_j} \right) - \frac{(\sum \sum y_{ij})^2}{n} \\ SSBT &= \frac{(45)^2}{3} + \frac{(47)^2}{3} + \dots + \frac{(21)^2}{3} - \frac{(350)^2}{27} \\ &= 4761.33 - 4537.0370 = 224.296. \end{aligned}$$

وبما أن عدد المجاميع 9 لذا فإن لها درجات حرية تساوي 8 .

مجموع المربعات الخطأ ( البواقي ) سيكون :

$$SSE = SYY - SSBT$$

$$= 246.963 - 224.296 = 22.6667.$$

تلخص النتائج في جدول (١٢-٤) .

جدول (١٢-٤)

S.O.V	df	SS	MS	F
بين المجاميع	8	224.296	28.037	22.2647
الخطأ	18	22.6667	1.259.23	
الكلية	26	246.963		

والآن تجزأ مجموع المربعات للمجاميع إلى :

(١) مجموع مربعات الإنحدار لـ  $\beta_1, \beta_2$  أي :

$$\begin{aligned} SSR(\beta_1, \beta_2 | \beta_0) &= b'X'y - \frac{(350)^2}{27} \\ &= 4715.09 - 4537.0370 \\ &= 178.053 . \end{aligned}$$

(٢) مجموع المربعات لنقص التوفيق سيكون :

$$\begin{aligned} SS(LF) &= SSBT - SSR(\beta_1, \beta_2 | \beta_0) \\ &= 224.296 - 178.053 \\ &= 46.243 . \end{aligned}$$

جدول تحليل التباين معطى في جدول (٤-١٣)

جدول (٤-١٣)

S.O.V	df	SS	MS	F
بين المجاميع	8	224.296		
$\beta_1, \beta_2   \beta_0$	2	178.053		
نقص التوفيق	6	46.243	7.7072	
الخطأ	18	22.666	1.25922	6.1206
الكلي	26	246.963		

وبما أن  $F$  المحسوبة (6.1206) تزيد عن قيمة  $F$  الجدولية  $F_{0.05}[6,18]=2.66$  عند مستوى معنوية  $\alpha = 0.05$  فإننا نرفض فرض العدم  $H_0$  أن النموذج خطي .

(١٠-٤) مشكلة عدم تجانس الخطأ ومعالجتها

نمذج الفروض على نموذج الانحدار الخطي المتعدد أن :

$$E(\varepsilon) = 0, \text{Cov}(\varepsilon) = \sigma^2 I .$$

في بعض الأحيان لا تتحقق تلك الفروض وهذا يؤدي إلى أن تقدير المربعات الصغرى بالطريقة العادية لا يمكن أن يكون أفضل تقدير خطي غير متحيز كما أن ثباين B يكون متحيزاً وبالتالي عملية اختبارات الفروض وفترات الثقة تكون غير صحيحة .

في الانحدار الخطي البسيط استخدم اختبار جولد فيلد - كواندت وذلك للكشف عن عدم تجانس الثباين حيث ترتب البيانات للمتغير x تصاعدياً ويتم تقسيمها إلى قسمين مع حذف بعض المشاهدات الوسطى وإيجاد الإحصاء F الذي يستخدم في الاختبار . أما في حالة الانحدار الخطي المتعدد ووجود عدة متغيرات مستقلة فإنه يكون من الصعب ترتيب البيانات تصاعدياً ولكن عادةً ينتخب أحد المتغيرات المستقلة التي يوجد نمط معين بين قيمة وقيم  $\varepsilon_j$  وترتب القيم تصاعدياً لقيم y وليقية قيم المتغيرات المستقلة بناءً على ذلك المتغير المنتخب وتقسّم البيانات إلى قسمين مع حذف بعض القيم الوسطى وعمل انحدار متعدد لكل قسم ثم إيجاد MSE لكل قسم واختبار تساويهما باستخدام الإحصاء :

$$F = \frac{MSE_1}{MSE_2} .$$

إذا كانت قيمة F المحسوبة من الإحصاء F تزيد عن القيمة الجدولية والمستخرجة من جدول توزيع F عند درجات حرية الخاصة بـ  $MSE_1$  و  $MSE_2$  فإننا نرفض فرض العدم، أي عدم تجانس الثباين .

معالجة عدم تجانس الثباين:

في هذا الجزء سوف نتناول طريقة المربعات الصغرى المرجحة في حالة الانحدار الخطي المتعدد عندما يكون  $\text{Cov}(\varepsilon) = \sigma^2 \Sigma$  حيث  $\Sigma$  مصفوفة من الدرجة  $n \times n$ . هذه الحالة سوف تكون سهلة في التفسير عندما تكون المصفوفة  $\Sigma$  مصفوفة قطرية ولكن بعناصر غير متساوية على القطر. وعلى ذلك المتغيرات

$y_1, y_2, \dots, y_n$  تكون غير مرتبطة ولكن لهم تباين غير متساوي ، بينما إذا كان بعض العناصر الغير قطرية للمصفوفة  $\Sigma$  غير صفرية فهذا يعني أن  $y_1, y_2, \dots, y_n$  سوف تكون مرتبطة .  
عندما يكون النموذج على الشكل التالي :

$$Y = XB + \varepsilon,$$

$$E(\varepsilon) = 0, \text{Cov}(\varepsilon) = \sigma^2 \Sigma$$

فإن تقديرات المربعات الصغرى العادية سوف تكون:

$$b = (X'X)^{-1} X'y$$

ولن تكون ملائمة . وفي هذه الحالة سوف نحول النموذج إلى فئة من المشاهدات التي تحقق فروض المربعات الصغرى العادية ثم استخدام طريقة المربعات الصغرى العادية على البيانات المحولة. وبما أن  $\sigma^2 \Sigma$  تمثل مصفوفة التغاير للأخطاء فإن  $\Sigma$  لابد أن تكون غير شاذة وكيدة الإيجابية وعلى ذلك يوجد مصفوفة  $K$  من الدرجة  $n \times n$  غير شاذة ومتماثلة حيث  $KK' = KK = \Sigma$ .  
المصفوفة  $K$  عادة تسمى مصفوفة الجذر التربيعي للمصفوفة  $\Sigma$ .  
سوف نعرف متغيرات جديدة:

$$Z = K^{-1}Y, U = K^{-1}X, g = K^{-1}\varepsilon.$$

على ذلك نموذج الانحدار :  $Y = XB + \varepsilon$  سوف يصبح :  
 $Z = U\beta + g$  أو  $K^{-1}Y = K^{-1}X\beta + K^{-1}\varepsilon$

$$Z = U\beta + g. \quad (٧-٤)$$

الأخطاء في التوزيع المحول لها توقع صفر . أي أن  $E(g) = K^{-1}E(\varepsilon) = 0$   
وأكثر من ذلك مصفوفة التغاير لـ  $g$  سوف تكون :

$$\text{Cov}(g) = \{[g - E(g)] [g - E(g)]'\}$$

$$\begin{aligned}
 &= E(gg') \\
 &= E(K^{-1}\varepsilon\varepsilon'K^{-1}) \\
 &= K^{-1}E(\varepsilon\varepsilon')K^{-1} \\
 &= \sigma^2 K^{-1}\Sigma K^{-1} \\
 &= \sigma^2 K^{-1}KKK^{-1} \\
 &= \sigma^2 I.
 \end{aligned}$$

وعلى ذلك عناصر  $g$  له متوسط صفر وتباين ثابت وغير مرتبطين . وعلى ذلك الأخطاء في النموذج (٧-٤) تحقق الفروض العادية عند تطبيق طريقة المربعات العادية .

معادلات المربعات الصغرى الطبيعية سوف تكون :

$$(X'\Sigma^{-1}X) b = X'\Sigma^{-1}y \quad (٨-٤)$$

والحل لتلك المعادلات سوف يكون :

$$b = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}y$$

حيث  $b$  يسمى تقدير المربعات الصغرى المرجح للمعلمة  $\beta$  .

المقدر  $B$  سوف يكون غير متحيز للمعلمة  $\beta$  وأيضا مصفوفة التغاير سوف تكون :

$$\text{Cov}(B) = \sigma^2 (U'U)^{-1} = \sigma^2 (X'\Sigma^{-1}X)^{-1}.$$

أيضا فإن  $B$  يمثل أفضل مقدر خطي للمعلمة  $\beta$ . جدول تحليل التباين في حالة استخدام المربعات المرجحة موضح في جدول (١٤-٤)

جدول (١٤-٤)

S.O.V	df	SS	MS	F
الإنحدار	p	SSR = $b'U'z$	SSR/p	$\frac{MSR}{MSE}$
الخطأ	n-p	SSE = $z'z - b'U'z$	SSE/(n-p)	
الكلية	n	$z'z$		



هناك صيغ أخرى لمجموع المربعات حيث:

مجموع مربعات الانحدار يساوي:

$$SSR = b'X'\Sigma^{-1}y$$

بدرجات حرية  $p$  . ومجموع المربعات الكلي يساوي:

$$SYY = y'\Sigma^{-1}y$$

بدرجات حرية  $n$  . ومجموع مربعات البواقي يساوي:

$$SSE = y'\Sigma^{-1}y - b'X'\Sigma^{-1}y$$

بدرجات حرية  $n-p$  . متوسط مجموع المربعات للانحدار هو:

$$MSR = \frac{SSR}{p}$$

ومتوسط مجموع مربعات البواقي هو:  $MSE = \frac{SSE}{(n-p)}$

عندما تكون الأخطاء  $\varepsilon$  غير مرتبطة ولكن لها تباين غير متساوي فإن مصفوفة التغاير تأخذ الصيغة التالية :

$$\sigma^2 \Sigma = \sigma^2 \begin{bmatrix} \frac{1}{w_1} & & & 0 \\ & \frac{1}{w_2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{w_n} \end{bmatrix}$$

العناصر  $w_i$  تسمى الأوزان.

المصفوفة  $\Sigma^{-1}$  سيوف تكون على الشكل :

$$W = \Sigma^{-1} = \begin{bmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n \end{bmatrix}$$

ومن (٤-٨) فإن المعادلات الطبيعية للمربعات الصغرى المرجحة سوف تكون :

$$(X'WX)b = X'Wy,$$

و:

$$b = (X'WX)^{-1} X'Wy,$$

تمثل تقديرات المربعات الصغرى المرجحة .

يمكن الحصول على تقديرات المربعات الصغرى المرجحة بسهولة باستخدام برنامج على الحاسب الآلي يحسب تقديرات المربعات الصغرى العادية وذلك بضرب كل مشاهدة  $x_{ij}$  (بما فيها العمود الأول في المصفوفة  $X$  والذي عناصره الواحد الصحيح) بالجذر التربيعي للوزن الخاص بتلك المشاهدة وبذلك نحصل على فئة من البيانات المعولة بحيث أن :

$$U = \begin{bmatrix} 1\sqrt{w_1} & x_{11}\sqrt{w_1} & \dots & x_{k1}\sqrt{w_1} \\ 1\sqrt{w_2} & x_{12}\sqrt{w_1} & \dots & x_{k2}\sqrt{w_2} \\ \vdots & \vdots & & \vdots \\ 1\sqrt{w_n} & x_{1n}\sqrt{w_1} & \dots & x_{kn}\sqrt{w_n} \end{bmatrix}$$

$$Z = \begin{bmatrix} y_1\sqrt{w_1} \\ y_2\sqrt{w_2} \\ \vdots \\ y_n\sqrt{w_n} \end{bmatrix}$$

والآن عند تطبيق طريقة المربعات الصغرى العادية على تلك البيانات المعولة نحصل على تقديرات المربعات الصغرى التالية :

$$b = (U'U)^{-1} U'Z = (X'WX)^{-1} X' Wy$$

أي تقديرات المربعات الصغرى المرجحة .

لإستخدام طريقة المربعات الصغرى المرجحة لابد أن تكون  $w_i$  معلومة. في بعض الأحيان فإن المعلومات المبدئية أو المعلومات عن النموذج يمكن استخدامها في تقدير الأوزان كما أوضحنا في البند (٢-٤) عند تناولنا طريقة المربعات الصغرى المرجحة في حالة نموذج الإتحاد الخطي البسيط. فعلى سبيل المثال عندما  $\text{Var}(e_j) = \sigma^2 x_{ij}$  فإن  $w_j = 1/x_{ij}$ .

مثال (٦-٤)

يعطى جدول (١٥-٤) بيانات لعينة حجمها  $n=5$  وذلك لمتغير مستقل ومتغير تابع وذلك تحت فرض النموذج الخطي التالي:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

$$\varepsilon_i \sim N(0, \sigma^2 x_i^2)$$

أوجد:

(أ) تقدير معالم النموذج.

(ب) تقدير مصفوفة التباين لمعالم هذا التوزيع.

جدول (١٥-٤)

x	1	2	3	4	5
y	3	8	5	6	8

الحل

الأوزان هنا سوف تكون  $w_i = 1/x_i^2$  وعلى ذلك المصفوفة  $\Sigma$  تأخذ الشكل التالي:

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 0 & 25 \end{bmatrix}, X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix},$$

$$W = \Sigma^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{9} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{16} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{25} \end{bmatrix}.$$

وبما أن :

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (X'WX)^{-1} X'Wy.$$

فإن :

$$(X'WX)^{-1} = \left( \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{9} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{16} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{25} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \right)^{-1}.$$

أذن :

$$(X'WX)^{-1} = \begin{bmatrix} 1.46361 & 2.28333 \\ 2.28333 & 5 \end{bmatrix}^{-1}.$$

وكذلك :

$$X'Wy = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 & 0 \\ 0 & 0 & 0.111 & 0 & 0 \\ 0 & 0 & 0 & 0.0625 & 0 \\ 0 & 0 & 0 & 0 & 0.04 \end{bmatrix} \begin{bmatrix} 3 \\ 8 \\ 5 \\ 6 \\ 8 \end{bmatrix}.$$

أذن:

$$X'Wy = \begin{bmatrix} 6.25056 \\ 11.7667 \end{bmatrix}.$$

وعلى ذلك :

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 2.37592 & -1.08501 \\ -1.08501 & 0.695486 \end{bmatrix} \begin{bmatrix} 6.25056 \\ 11.7667 \end{bmatrix} = \begin{bmatrix} 2.08395 \\ 1.40166 \end{bmatrix}.$$

معادلة الإحداد المقدرة ستكون:

$$\hat{y} = 2.08395 + 1.40166x.$$

التقدير لمصفوفة التباين - التغاير للمقدر B هو:

$$Cov(B) = s^2 (X'WX)^{-1}.$$

حيث أن:

$$s^2 = \frac{y'Wy - b'X'Wy}{n - p}.$$

$$y'Wy = \begin{bmatrix} 3 & 8 & 5 & 6 & 8 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 & 0 \\ 0 & 0 & 0.111 & 0 & 0 \\ 0 & 0 & 0 & 0.0625 & 0 \\ 0 & 0 & 0 & 0 & 0.04 \end{bmatrix} \begin{bmatrix} 3 \\ 8 \\ 5 \\ 6 \\ 8 \end{bmatrix}.$$

إذن:

$$y'Wy = 32.5878,$$

$$b'X'Wy = \begin{bmatrix} 2.08395 & 1.40166 \end{bmatrix} \begin{bmatrix} 6.25056 \\ 11.7667 \end{bmatrix} = 29.5187.$$

الآن:

$$s^2 = (32.5876 - 29.5187) / 3 = 1.02301.$$

الآن:

$$\begin{aligned} \text{Cov}(B) &= 1.02301 \begin{bmatrix} 2.37592 & -1.08501 \\ -1.08501 & 0.695486 \end{bmatrix} \\ &= \begin{bmatrix} 2.4306 & -1.10997 \\ -1.10997 & 0.71149 \end{bmatrix} \end{aligned}$$

وعلى ذلك تبين  $B_0, B_1$  المقدر هما:

$$S^2(b_0) = 2.4306, S^2(b_1) = 0.71149,$$

التغاير بين  $B_0, B_1$  المقدر هو:

$$S(b_0, b_1) = -1.10997.$$

## الفصل الخامس

### اختيار أفضل نموذج إحدار

### Selecting the Best Regression Model

مقدمه	(١-٥)
معامل التحديد المتعدد	(٢-٥)
إحصاء ملاوس $C_p$	(٣-٥)
متوسط مجموع مربعات البواقي	(٤-٥)
المقياس $PRESS_p$	(٥-٥)
طريقة كل الانحدارات الممكنة	(٦-٥)
طريقة الحذف الخلفي (العكسي)	(٧-٥)
طريقة الاختيار الامامي (المباشر)	(٨-٥)
طريقة الاختيار التكريري	(٩-٥)

## (١-٥) مقدمه

يتناول هذا الفصل بعض الأساليب الإحصائية لاختيار أفضل نموذج إنحدار للمتغير التابع ومجموعة المتغيرات المستقلة  $x_1, x_2, \dots, x_k$  حيث يتم اختيار المتغيرات المستقلة الأكثر تأثيراً على المتغير التابع  $Y$ . وهناك بعض الآراء في هذا المجال منها:

١- محاولة إدخال أكبر عدد ممكن من المتغيرات المستقلة في نموذج الإنحدار حتى تكون القيم المتنبأ بها للمتغير التابع  $Y$  أكثر دقة.

٢- محاولة إدخال أقل عدد ممكن من المتغيرات المستقلة في نموذج الإنحدار حيث أن الحصول على معلومات عن عدد كبير من المتغيرات قد يكون أكثر تكلفة.

وبين هذه الآراء يكون موضوع " اختيار أفضل نموذج إنحدار " . وسنعرض الآن بعض الأساليب التي تساعد في هذا الاختيار مع ملاحظة أن هذه الأساليب قد لاتعطي نفس النتائج بالنسبة لمشكلة معينة.

أ- طريقة كل الانحدارات الممكنة.

All possible regressions

ب- طريقة الحذف الخلفي (العكسي).

The backward elimination procedure

ج- طريقة الاختيار الامامي (المباشر).

The forward selection procedure

د- طريقة الاختيار التدريجي .

The stepwise selection procedure .

هذا وسنستخدم عدة مقاييس للمفاضلة بين المعادلات المرشحة وقيل القيام بذلك ، نحتاج إلى اعتماد بعض الرموز فلنرمز لعدد المتغيرات المستقلة المرشحة في الجملة بـ  $P-1$  ونفترض ، في هذا الفصل ، أن جميع نماذج الإنحدار تتضمن الجزء المقطوع  $\beta_0$ . سوف نرمز لعدد المتغيرات المستقلة في مجموعة جزئية بـ  $p-1$  ، اي يوجد  $p$  معلمه في نموذج الإنحدار الخاص بهذه المجموعة الجزئية من المتغيرات المستقلة (النموذج المخفض). وعلى ذلك فإن:

$$1 \leq p \leq P.$$

ويفرض ان عدد المشاهدات  $n$  اكبر من عدد المعالم المرشحة أي  $n > P$ .



ومن المستحسن جداً أن يكون  $n$  أكبر بكثير من  $P$  بحيث يمكن الحصول على نتائج سليمة.

### (٢-٥) معامل التحديد المتعدد

ليكن  $R_p^2$  هو معامل التحديد لنموذج الانحدار المخفض بحدود عددها  $p$ ، أي يوجد  $p-1$  من المتغيرات والجزء المقطوع  $\beta_0$  حسابياً فإن:

$$R_p^2 = \frac{SSR(p)}{SYY} = 1 - \frac{SSE(p)}{SYY} \quad (١-٥)$$

حيث  $SSR(p)$ ،  $SSE(p)$  في (١-٥) يمثلان مجموع مربعات الانحدار ومجموع مربعات البواقي على التوالي للنموذج المخفض. وبما أن المقام ثابت في جميع نماذج الانحدار الممكنة فإن  $R_p^2$  يتغير عكسياً مع مجموع الخطأ  $SSE(p)$ . ومن المعروف أن  $R_p^2$  يكون أكبر ما يمكن عندما يحتوي نموذج الانحدار على جميع المتغيرات المستقلة المرشحة وعددها  $p-1$  وبالتالي لا يمكن أن يكون سبب استخدام المقاييس  $R_p^2$  عند اختيار أفضل نموذج إنحدار هو جعل  $R_p^2$  أعظم ما يمكن. وإنما الهدف هو إيجاد الوضع الذي تصبح إضافة المزيد من المتغيرات المستقلة عنده غير ذات شأن بإعتبارها تؤدي إلى زيادة صغيرة جداً في  $R_p^2$ . وفي الغالب نصل إلى هذا الوضع عندما يحتوي نموذج الانحدار على عدد محدود فقط من المتغيرات المستقلة. ومن الواضح، أن تحديد الوضع الذي يبدأ عنده العائد بالتناقص هي مسألة اجتهاد. لقد اقترح Aitkin (1974) طريقة لاختيار نماذج الانحدار الجزئية والتي لها  $R^2$  غير معنوي عن  $R^2$  الخاصة بالنموذج الكامل. ليكن:

$$R_0^2 = 1 - (1 - R_p^2)(1 + d_{\alpha, n, p-1})$$

حيث:

$$d_{\alpha, n, p-1} = (P-1) \frac{F_{\alpha}(n, n-P)}{n-P}$$

حيث  $R_p^2$  قيمة معامل التحديد المتعدد للنموذج الكامل. ويعبر Aitkin أي فئة من المتغيرات المستقلة تعطي  $R^2$  أكبر من  $R_0^2$  فئة مقابله. عادة يرسم  $R_p^2$  مقابل  $p$  للمساعدة في اختيار أفضل نموذج إنحدار.

### معامل التحديد المعدل

لتجنب الصعوبات في تفسير  $R^2$  ، فإن بعض الباحثين يفضلون استخدام معامل التحديد المعدل [(Ezekiel (1930)] والمعرف لحدود عددها  $p$  في نموذج الانحدار المخفض كالتالي:

$$\bar{R}_p^2 = 1 - \frac{(n-1)}{n-p} (1 - R_p^2).$$

عادة يختار النموذج الذي له القيمة العظمى لـ  $\bar{R}_p^2$ .

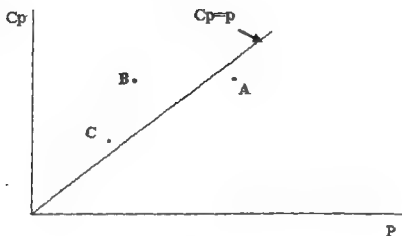
### (٣-٥) إحصاء ملاوس (Mallows' $C_p$ Statistic)

يلخذ إحصاء ملاوس  $C_p$  الصيغة التالية:

$$C_p = \frac{SSE(p)}{MSE} - [n - 2p]$$

حيث  $SSE(p)$  مجموع مربعات البواقي للنموذج المخفض و  $MSE$  متوسط مجموع مربعات للنموذج الكامل. ويساعد إحصاء ملاوس في تحديد عدد المتغيرات المستقلة التي يجب إدخالها في نموذج الانحدار الأفضل وذلك لأن قيمة إحصاء ملاوس يساوي تقريباً  $p$  حيث  $C_p = p$  ، عدد المعالم في النموذج المخفض وذلك عندما يكون تباين النموذج المخفض يساوي تقريباً تباين النموذج الكامل [(Mallows (1973)]. فالنموذج الجيد هو الذي يساوي عدد معاملاته  $p$  قيمة إحصاء ملاوس ويستخدم انحراف قيمة  $C_p$  عن عدد معالم النموذج كمقياس للتحيز.

عند استخدام  $C_p$  كمقياس للمفاضلة بين النماذج المرشحة نرسم  $C_p$  مقابل  $p$  لكل نموذج انحدار كما هو موضح في شكل (١-٥). نماذج الانحدار التي تقع قريبه من الخط  $C_p = p$  يكون لها تحيز قليل بينما النماذج التي تقع فوق الخط (مثل النقطة B في شكل (١-٥) ) يكون لها تحيز كبير. عموماً يفضل القيم الصغيرة من  $C_p$  . على سبيل المثال النقطة C في شكل (١-٥) فوق الخط  $C_p = p$  واسفل النقطة A والتي تمثل نموذج له أقل مجموع أخطاء.



شكل (١-٥)

#### (٤-٥) متوسط مجموع مربعات البواقي

يحسب متوسط مجموع مربعات البواقي لنموذج الانحدار المخفض من الصيغة التالية:

$$MSE(p) = \frac{SSE(p)}{n - p}$$

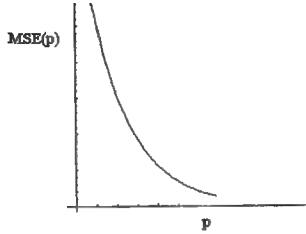
والذي يستخدم كمقياس للمفاضلة بين النماذج المرشحة. عند رسم  $MSE(p)$  مقابل  $p$  كما هو موضح في شكل (٢-٥) نجد أنه في البدايه تتناقص  $MSE(p)$  مع زيادة  $p$  ثم تثبت وتندثر ما تريد . وعادة اختيار  $p$  من الرسم يتوقف على:

(١) أقل قيمة لـ  $MSE(p)$ .

(٢) القيمة لـ  $p$  بحيث أن  $MSE(p)$  تقريباً مساوية لـ  $MSE$  للنموذج الكامل.

النماذج الجزئية التي تؤدي الى تصغير  $MSE(p)$  ايضاً تؤدي الى تعظيم

$$\bar{R}_p^2$$



شكل (٢-٥)

لأثبات ذلك فإن:

$$\begin{aligned}\bar{R}_p^2 &= 1 - \frac{n-1}{n-p} (1 - \bar{R}_p^2) \\ &= 1 - \frac{n-1}{n-p} \frac{SSE(p)}{SYY} \\ &= 1 - \frac{n-1}{SYY} \frac{SSE(p)}{n-p} \\ &= 1 - \frac{n-1}{SYY} MSE(p)\end{aligned}$$

وعلى ذلك فإن تصغير  $MSE(p)$  يكافئ تعظيم  $\bar{R}_p^2$ .

#### ٥-٥) المقياس $PRESS_p$

يعتمد المقياس  $PRESS_p$  (مجموع مربعات التنبؤ) على بواقى الحذف  $d_i$  المعرفه كالتالي:

$$d_j = y_j - \hat{y}_{(j)}$$

حيث  $\hat{y}_{(j)}$  هي القيمة المقدرة للمشاهدة  $j$  عند تقدير نموذج الانحدار بعد حذف المشاهدة  $j$ . ولكل نموذج انحدار  $n$  من بواقي الحذف المرافقة له. والمقياس  $PRESS_p$  يمثل مجموع مربعات بواقي الحذف:

$$PRESS_p = \sum_{j=1}^n d_j^2 = \sum_{j=1}^n (y_j - \hat{y}_{(j)})^2$$

وتعتبر النماذج ذات القيمة الصغيرة للمقياس  $PRESS_p$  نماذج جيدة. ويمكن حساب قيمة  $PRESS_p$  من المعادلة التالية:

$$PRESS_p = \sum_{j=1}^n \left( \frac{e_j}{1 - h_{jj}} \right)^2$$

حيث  $e_j$  هو الباقي و  $h_{jj}$  هو العنصر رقم  $j$  على القطر للمصفوفة  $H$ .

#### (٦-٥) طريقة كل الانحدارات الممكنة

يعتبر هذا الأسلوب من الأساليب التي تتطلب العديد من العمليات الحسابية التي يصعب إجرائها دون استخدام الحاسبات الآلية. وسوف نشرح هذه الطريقة بالمثل التالي والذي سوف يقتصر في الحساب على المقاييس الثلاثة الأولى فقط وذلك لأن المقياس الرابع ( $PRESS_p$ ) يحتاج إلى عدد كبير من العمليات الحسابية الخاصة بالانحدار وهو متوفر في كثير من حزم الحاسب الآلي الجاهزة .

#### مثال (١-٥)

بفرض أن لدينا عينه من الحجم  $n = 17$  من المشاهدات لقيم متغير الاستجابة  $Y$  مع متغيرات مستقلة مرشحة عددها  $P-1 = 5$  والمعطاه في جدول (١-٥).

جدول (١-٥)

رقم المشاهد	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$y$
1	58.80	7107.00	21.00	129.00	52.00	3067.00
2	65.20	6373.00	22.00	141.00	68.00	2828.00
3	70.90	6796.00	22.00	153.00	29.00	2891.00
4	77.40	9208.00	20.00	166.00	23.00	2994.00
5	79.30	14792.00	25.00	193.00	40.00	3082.00
6	81.00	14564.00	23.00	189.00	14.00	3898.00
7	71.90	11964.00	20.00	175.00	96.00	3502.00
8	63.90	13526.00	23.00	186.00	94.00	3060.00
9	54.50	12656.00	20.00	190.00	54.00	3211.00
10	39.50	14119.00	20.00	187.00	37.00	3286.00
11	44.50	16691.00	22.00	195.00	42.00	3542.00
12	43.60	14571.00	19.00	206.00	22.00	3125.00
13	56.00	13619.00	22.00	198.00	28.00	3022.00
14	46.70	14575.00	22.00	192.00	7.00	2922.00
15	73.00	14556.00	21.00	191.00	42.00	3950.00
16	78.90	18573.00	21.00	200.00	33.00	4488.00
17	79.40	15618.00	22.00	200.00	92.00	3295.00

ولاختيار أفضل المتغيرات المستقلة التي عددها  $p-1$  للمثال (١-٥) نتبع الخطوات التالية:

١- نقرر كل أنواع نماذج الانحدار الممكنة باستخدام كل المجاميع الممكنة من المتغيرات المستقلة. فلو كان لدينا  $P-1$  من المتغيرات المستقلة المرشحة فإن

عدد النماذج الكليه سيكون  $2^{P-1}$  حيث تشملها النموذج الذي لا يحتوي على متغيرات مستقلة ويحتوي فقط على  $\beta_0$ . وكل نموذج يحتوي على المعامل الثابت  $\beta_0$  وكلما زاد عدد المتغيرات المستقلة المرشحة في نموذج الانحدار كلما زاد عدد النماذج الممكن تقديرها بشكل سريع. فعلى سبيل المثال عند  $P = 10$  فإن عدد النماذج الممكن تقديرها تساوي  $2^{10} = 1024$ . الآن للمثال (٥-١) حيث عدد المتغيرات يساوي  $P-1 = 5$  فإن عدد النماذج الممكن تقديرها هي  $2^5 = 32$ . أن النماذج هذه يمكن وضعها في خمس مجاميع هي:

- مجموعة A وتمثل النموذج الذي لا يحتوي على أي متغير مستقل.
- مجموعة B وهي فئة النماذج التي تحتوي على متغير مستقل واحد.
- مجموعة C وهي فئة النماذج التي تحتوي على متغيرين مستقلين.
- مجموعة D وهي فئة النماذج التي تحتوي على ثلاثة متغيرات مستقلة.
- مجموعة E وهي فئة النماذج التي تحتوي على أربعة متغيرات مستقلة.
- مجموعة F وهي فئة النماذج التي تحتوي على خمسة متغيرات مستقلة.

إن نتائج التحليل الإحصائي لجميع النماذج الانحداريه اعلاه ملخصه في جدول (٥-٢). فمثلاً عندما تكون المتغيرات المستقلة  $x_1, x_2$  في المعادلة ، فإن المعادلة المقدرة سوف تكون:

$$\hat{y} = 1475.812 + 11.910x_1 + 0.083x_2.$$

أيضاً عندما تكون المتغيرات المستقلة  $x_1, x_4$  موجوده في النموذج فإن معادلة الانحدار المقدرة سوف تكون:

$$\hat{y} = 825.999 + 12.318x_1 + 9.305x_4.$$

وهكذا لكل نموذج في جدول (٥-٢).

٢- نصيب كل من  $R_p^2$  و  $C_p$ ,  $MSE(p)$  لكل معادلة فمثلاً للنموذج:

٢- نحسب كل من  $R_p^2$  و  $MSE(p)$  لكل معادلة فمثلا للنموذج:

$$Y_i = \beta_0 + \beta_2 x_{2j} + \epsilon_j$$

فلن:

$$R_p^2 = R_2^2 = \frac{SSR(2)}{SYY} = \frac{1270172.19}{3192631.53} = 0.398 ,$$

$$MSE(p) = MSE(2) = \frac{SSE(2)}{n-p} = \frac{1922459.34}{17-2}$$

$$= \frac{1922459.34}{15} = 128163.956,$$

$$\bar{R}_p^2 = 1 - \frac{(n-1)}{(n-p)} (1 - R_p^2) = 1 - \frac{16}{15} (1 - 0.398) = 0.358 ,$$

$$\begin{aligned} C_p = C_2 &= \frac{SSE(2)}{MSE} - (n-2p) \\ &= \frac{1922459.34}{46642.157} - (17 - (2)(2)) \\ &= 28.217. \end{aligned}$$

وهكذا وننظم النتائج كما هي معطاه في جدول (٢-٥).

ومثلا للنموذج :

$$Y_j = \beta_0 + \beta_1 x_j + \beta_4 x_{4j} + \epsilon_j$$

فلن:

$$R_p^2 = R_3^2 = \frac{SSR(3)}{SYY} = \frac{1029010.31}{3192631.53} = 0.322,$$



$$= 1 - \frac{16}{14}(1 - 0.322) = 0.225,$$

$$C_p = C_3 = \frac{SSE(3)}{MSE} - (n - 2p)$$

$$= \frac{2163621.22}{46642.775} - (17 - (2)(3))$$

$$= 35.3877.$$

وهكذا ، تتظم النتائج كما هو موضح في جدول (٣-٥)

جدول (٣-٥)

المتغيرات في النموذج	تقديرات المبيعات الصغرى					
	b <sub>0</sub>	b <sub>1</sub>	b <sub>2</sub>	b <sub>3</sub>	b <sub>4</sub>	b <sub>5</sub>
x <sub>1</sub>	2623.212	10.667				
x <sub>2</sub>	2273.088		7.989E-02			
x <sub>3</sub>	3886.105			-27.125		
x <sub>4</sub>	1777.723				8.392	
x <sub>5</sub>	3352.216					-1.067
x <sub>1</sub> x <sub>2</sub>	1475.812	11.910	0.083			
x <sub>1</sub> x <sub>3</sub>	4121.274	13.824		-79.152		
x <sub>1</sub> x <sub>4</sub>	825.999	12.318			9.305	
x <sub>1</sub> x <sub>5</sub>	2668.263	11.636				-2.350
x <sub>2</sub> x <sub>3</sub>	3277.103		0.082	-48.028		
x <sub>2</sub> x <sub>4</sub>	4600.805		0.203		-21.567	
x <sub>2</sub> x <sub>5</sub>	2268.723		0.080			0.077
x <sub>3</sub> x <sub>4</sub>	2435.165			-31.180	8.459	
x <sub>3</sub> x <sub>5</sub>	3916.796			-26.408		-1.014
x <sub>4</sub> x <sub>5</sub>	1777.607				1.427E-20	8.393
x <sub>1</sub> x <sub>2</sub> x <sub>3</sub>	3506.503	16.455	0.089	-111.692		
x <sub>1</sub> x <sub>2</sub> x <sub>4</sub>	3653.250	10.159	0.192		-19.089	

$X_1 X_2 X_5$	1513.587	12.417	0.082			-1.265
$X_1 X_3 X_4$	2482.200	16.086		-92.353	9.780	
$X_1 X_3 X_5$	4218.220	14.978		-81.687		-2.554
$X_1 X_4 X_5$	895.000	12.824			9.079	-1.325
$X_2 X_3 X_4$	6306.807		0.218	-71.385	-23.547	
$X_2 X_3 X_5$	3269.490		0.082	-48.221		0.206
$X_2 X_4 X_5$	4699.172		0.205		-21.989	-0.932
$X_3 X_4 X_5$	2430.357			-31.240	8.474	0.074
$X_1 X_2 X_3 X_4$	6244.042	15.148	0.212	-127.881	-21.418	
$X_1 X_2 X_3 X_5$	3569.719	17.087	0.088	-112.756		-1.469
$X_1 X_2 X_4 X_5$	3796.485	10.897	0.194		-19.818	-2.011
$X_1 X_3 X_4 X_5$	2581.523	16.709		-93.511	9.529	-1.508
$X_2 X_3 X_4 X_5$	6384.489		0.219	-70.965	-23.912	-0.83
$X_1 X_2 X_3 X_4 X_5$	6458.748	16.100	0.215	-130.251	-22.310	-2.340

جدول (٢-٥)

المتغيرات في المعادلة	p	$R_p^2$	$\bar{R}_p^2$	MSE(p)	$C_p$
لا يوجد	1				
$X_1$	2	0.115	0.056	188334.103	47.5677
$X_2$	2	0.398	0.358	128163.956	28.21718
$X_3$	2	0.008	-0.058	211162.772	54.909388
$X_4$	2	0.171	0.115	176494.556	43.7821
$X_5$	2	0.004	-0.062	211925.5	55.1547
$X_1 X_2$	3	0.541	0.475	104701.456	20.4269
$X_1 X_3$	3	0.172	0.054	188765.7	45.6594
$X_1 X_4$	3	0.322	0.225	154544.373	35.3877
$X_1 X_5$	3	0.135	0.012	197239.562	48.2029
$X_2 X_3$	3	0.422	0.340	131740.898	28.5430

$X_2X_4$	3	0.574	0.513	97097.215	18.144
$X_2X_5$	3	0.398	0.312	137313.469	30.0157
$X_3X_4$	3	0.181	0.064	186726.287	45.0473
$X_3X_5$	3	0.012	-0.129	225360.131	56.6435
$X_4X_5$	3	0.171	0.052	189101.308	45.7602
$X_1X_2X_3$	4	0.652	0.572	85383.807	14.7979
$X_1X_2X_4$	4	0.676	0.601	79584.489	13.1816
$X_1X_2X_5$	4	0.547	0.442	111349.909	22.0352
$X_1X_3X_4$	4	0.400	0.261	147474.93	32.1039
$X_1X_3X_5$	4	0.196	0.010	197514.037	46.0507
$X_1X_4X_5$	4	0.329	0.174	164905.085	36.962
$X_2X_3X_4$	4	0.627	0.541	91661.579	16.5477
$X_2X_3X_5$	4	0.422	0.289	141836.115	30.5322
$X_2X_4X_5$	4	0.577	0.480	103784.974	19.9267
$X_3X_4X_5$	4	0.181	-0.008	201084.846	47.0459
$X_1X_2X_3X_4$	5	0.820	0.760	47879.996	5.318
$X_1X_2X_3X_5$	5	0.660	0.547	90448.368	16.27
$X_1X_2X_4X_5$	5	0.690	0.587	82423.668	14.2058
$X_1X_3X_4X_5$	5	0.408	0.210	157623.305	33.553
$X_2X_3X_4X_5$	5	0.629	0.506	98627.556	18.3811
$X_1X_2X_3X_4X_5$	6	0.839	0.766	46642.175	6.000

نتائج النتائج كالتالي:

(أ) عند استخدام  $R_p^2$  مقياساً للمفاضلة:

من جدول (٥-٣) نختار من كل فئة النموذج الذي يحتوي على أعلى  $R_p^2$  ونرتب النتائج كما هو موضح في جدول (٥-٤).

جدول (٤-٥)

	p	المتغيرات في النموذج	$R_p^2$
B	2	$x_2$	0.398
C	3	$x_2, x_4$	0.574
D	4	$x_1, x_2, x_4$	0.676
E	5	$x_1, x_2, x_3, x_4$	0.820
F	6	$x_1, x_2, x_3, x_4, x_5$	0.839

نقارن  $R_p^2$  للنماذج المرشحة في كل مجموعة. من جدول (٤-٥) نلاحظ أن الزيادة في قيمة  $R^2$  طفيفة جداً بين المجموعة E والمجموعة F أي أن إضافة متغير مستقل آخر في معادلة الانحدار في المجموعة E لن يؤدي إلى زيادة محسوسة في مجموع المربعات للانحدار وبالتالي لن يؤدي إلى زيادة محسوسة في  $R_p^2$  ومن ذلك يتضح أن أفضل معادلة إنحدار في هذه الحالة هي:

$$\hat{y} = 6244.42 + 15.148x_1 + 0.212x_2 - 127.881x_3 - 21.418x_4.$$

(ب) عند استخدام  $\bar{R}_p^2$  مقياساً للمفاضلة:

من جدول (٣-٥) نختار من كل فئة النموذج الذي يحتوي على أعلى  $\bar{R}_p^2$  وترتيب النتائج كما هو موضح في جدول (٥-٥).

جدول (٥-٥)

	p	المتغيرات في النموذج	$\bar{R}_p^2$
B	2	$x_2$	0.358
C	3	$x_2, x_4$	0.513
D	4	$x_1, x_2, x_4$	0.601
E	5	$x_1, x_2, x_3, x_4$	0.760
F	6	$x_1, x_2, x_3, x_4, x_5$	0.766

نقارن  $\bar{R}_p^2$  للنماذج المنتخبة من كل فئة. عند مقارنة  $\bar{R}_p^2$  للنماذج المرشحة من كل فئة فإننا يجب أن نختار النموذج في الفئة E لأن  $\bar{R}_p^2 = 0.76$  وذلك لأن إضافة  $x_5$  على النموذج الذي يحتوي على  $x_1, x_2, x_3, x_4$  له تأثير ضئيل جداً على الاستجابة  $Y$ .

(ج) عند استخدام  $C_p$  مقياساً للمفاضلة بين النماذج:  
من جدول (٣-٥) نختار من كل فئة النموذج الذي يحتوي على أقل  $C_p$  ونرتب النتائج كما هو موضح في جدول (٦-٥).

جدول (٦-٥)

الفئة	المتغيرات المستقلة في النموذج	$C_p$
B	$x_2$	28.2172
C	$x_2, x_4$	18.144
D	$x_1, x_2, x_4$	13.1816
E	$x_1, x_2, x_3, x_4$	5.318
F	$x_1, x_2, x_3, x_4, x_5$	6.0

من جدول (٦-٥) نجد أن أفضل نموذج هو النموذج الذي يحتوي على أربعة متغيرات مستقلة وينتمي إلى الفئة E. حيث  $C_p$  لهذا النموذج يساوي 5.318 وهو قريب من  $p = 5$  حيث  $p = 5$ .

(د) عند استخدام  $MSE(p)$  مقياساً للمفاضلة:  
من جدول (٣-٥) نختار من كل فئة النموذج الذي يحتوي على أقل  $MSE(p)$  ونرتب النتائج كما هو موضح في جدول (٧-٥).

جدول (٧-٥)

	p	المتغيرات في النموذج	MSE(p)
B	2	$x_2$	128163.956
C	3	$x_2, x_4$	97097.215
D	4	$x_1, x_2, x_4$	79584.489
E	5	$x_1, x_2, x_3, x_4$	47879.996
F	6	$x_1, x_2, x_3, x_4, x_5$	46642.175

فعلد مقارنة MSE(p) للنماذج المرشحة من كل فئة نرى أن أقلهم النموذج في الفئة F والذي يحتوي على خمسة متغيرات مستقلة. وهذا يفترض مع ماحصلنا عليه عند استخدام  $R_p^2$  و  $\bar{R}_p^2$  و  $C_p$ . وعلى كل حال فأحياناً كل مقياس من مقاييس المفاضلة قد يعطي نموذج مفضل يختلف بعضها عن البعض الآخر. في هذه الحالة فإن المفاضلة بين نتائج المقاييس يمكن أن يتم من خلال استخدام اختبار  $t$ . لمثالنا (مثال (١-٥)) فإن قيم  $t$  معطاه في جدول (٨-٥) وذلك لمعاملات الانحدار في النموذج الذي به 5 متغيرات و جدول (٩-٥) للنموذج الذي به أربعة متغيرات).

جدول (٨-٥)

المتغيرات المستقلة في النموذج	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
$t$	5.791	5.436	-3.195	-3.503	-1.148
المعنوية	.003	0.000	0.009	0.005	0.275

يتضح من صف المعنوية في جدول (٨-٥) والمستخرج من برنامج SPSS أن  $|t|$  الخاصة بالمتغير  $x_5$  هي الوحيدة الغير معنوية حيث  $0.275 > 0.05$  و  $\alpha = 0.05$  من جدول (٩-٥) يتضح معنوية كل معاملات الانحدار الخاصة بالمتغيرات المستقلة للنموذج الذي به أربعة متغيرات مستقلة والذي ينتمي إلى الفئة E. وهذا يعطي أن المفاضلة سوف تكون لهذا النموذج.

جدول (٩-٥)

المتغيرات المستقلة في النموذج	$x_1$	$x_2$	$x_3$	$x_4$
$t$	3.590	5.295	-3.100	-3.344
المعنوية	0.004	0.000	0.009	0.006

(٧-٥) طريقة الحذف الخلفي (العكسي)

تعتبر هذه الطريقة تحسين للطريقة السابقة حيث لا تسمح بفحص كل نماذج الانحدار الممكنة بل أفضل النماذج الذي تحتوي على عدد معين من المتغيرات المرشحة. وتتلخص هذه الطريقة بمايلي:

يبدأ النموذج بوجود جميع المتغيرات المستقلة المرشحة ثم تحذف المتغيرات المستقلة من النموذج واحد بعد الآخر وتتوقف عن الحذف عندما تكون قيمة  $F$  الجزئية أكبر من قيمة  $F$  الجدولية المعينة. وأول المتغيرات الذي يحذف من النموذج هو المتغير الذي له أقل قيمة  $F$  جزئية وتكون أقل من قيمة  $F$  الجدولية (أما إذا كانت أقل قيمة  $F$  جزئية أكبر من قيمة  $F$  الجدولية فتتوقف عن الحذف وتكون بذلك جميع المتغيرات في النموذج هي متغيرات مهمة وذات تأثير معنوي على الاستجابة  $Y$ ). وبعد حذف المتغير الأول نحسب قيم  $F$  الجزئية لبقية المتغيرات المستقلة ونحذف المتغير الذي له  $F$  الجزئية التي تكون أقل من قيمة  $F$  الجدولية المعينة وهكذا لكل خطوة. وتتوقف عن الحذف في المرحلة التي تكون أقل قيمة  $F$  جزئية أكبر من قيمة  $F$  الجدولية المعينة لتلك المرحلة وبذلك يكون النموذج الذي يحوي على بقية المتغيرات المستقلة غير المحذوفة هو أفضل النماذج. وسوف نستخدم في مثالنا اختبار  $t$  بدلا من  $F$  الجزئية وذلك لأن كثير من برامج الحاسب الآلي الجاهزة والخاصة بالانحدار تعتمد عليه حيث  $F = t^2$ .

وسوف نطبق هذه الطريقة على مثال (١-٥) كالتالي:

الخطوة الأولى:

١- نوجد معادلة الانحدار المقدره على جميع المتغيرات المستقلة المرشحة وهي:

$$\hat{y} = 6458.748 + 16.100x_1 + 0.215x_2 - 130.252x_3 - 22.310x_4 - 2.340x_5$$

يتم معرفة مدى مساهمة كل متغير في معادلة الانحدار في مجموع المربعات للانحدار، ولتحديد هذه المساهمة للمثال (١-٥) يتم حساب قيمة  $t$  لكل متغير في النموذج حيث قيم  $t$  لكل متغير مستقل معطاه في جدول (١٠-٥).

جدول (١٠-٥)

المتغيرات المستقلة في النموذج	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
$t$	3.791	5.436	-3.195	-3.503	-1.148

وبما أن أقل قيمة لـ  $|t|$  هي للمتغير  $x_5$  حيث  $|t| = 1.148$  والتي نقل عن قيمة  $t$  الجدوليه حيث:

$$t_{0.025}(11) = 2.201.$$

ولذلك يتم حذف  $x_5$  من نموذج الانحدار. نوجد معادلة الانحدار المقدر للمتغيرات  $x_1, x_2, x_3, x_4$  وهي:

$$\hat{y} = 6244.042 + 15.148x_1 + 0.212x_2 - 127.881x_3 - 21.418x_4.$$

قيم  $t$  لكل متغير في النموذج معطاه في جدول (١١-٥).

جدول (١١-٥)

المتغيرات المستقلة في النموذج	$x_1$	$x_2$	$x_3$	$x_4$
$t$	3.590	5.295	-3.100	-3.344



وبما أن أقل قيمة لـ  $|t|$  هي للمتغير  $x_3$  والتي تزيد عن قيمة  $t$  الجدوليه حيث:

$$t_{0.025}(12) = 2.179.$$

أي أن مساهمة المتغير  $x_3$  فعالة في نموذج الانحدار ويجب الاحتفاظ به وبذلك تنتهي خطوات هذه الطريقة ويكون أفضل نموذج هو الذي يحتوي على المتغيرات  $x_1, x_2, x_3, x_4$  حيث معامل التحديد هنا هو  $R^2_5 = 0.82$  بينما معامل التحديد في حالة وجود  $x_5, x_4, x_3, x_2, x_1$  في معادلة الانحدار هو  $R^2_6 = 0.839$  حيث يلاحظ أن الفرق بسيط بينهما.

#### (٨-٥) طريقة الاختيار الاسمي (المباشر)

كانت طريقة الحذف الخلفيه تبدأ بتقدير نموذج الانحدار باستخدام جميع المتغيرات المستقلة وبالتدرج يتم حذف عدد من المتغيرات في النموذج حيث نصل الى قرار للنموذج الذي يتم استخدامه ، اما طريقة الاختيار الامامية فهي محاولة للحصول على نتيجة مشابهة ولكن العمل في الاتجاه الآخر وهو ان ندخل المتغيرات بالتدرج حتى نحصل على أفضل نموذج إحدار وطريقة الإدخال تتحدد باستخدام معامل الارتباط الجزئي وكلما أدخل متغير الى نموذج الانحدار يتم إجراء الآتي:

- تقدير معامل التحديد  $R^2_p$ .

- إجراء اختبار  $t$  لآخر متغير أدخل في نموذج الانحدار لمعرفة ما إذا كان هذا المتغير أضاف جزءاً معنوياً الى مجموع المربعات المفسرة أم لا ؟

وبمجرد أن نحصل على قيمة لـ  $|t|$  لآخر متغير أدخل في نموذج الانحدار غير معنوية يتم حذف هذا المتغير من نموذج الانحدار وتنتهي العملية عند ذلك

وستطبق هذه الطريقة على المثال (١-٥) كما يلي:

#### الخطوة الأولى:

١- تبدأ بإيجاد مصفوفة معاملات الارتباط البسيطة بين جميع المتغيرات المستقلة المرشحة ومتغير الاستجابة  $Y$  والتي تكون على الشكل التالي:

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$y$
$x_1$	1	$r_{12}$	$r_{13}$	$r_{14}$	$r_{15}$	$r_{y1}$
$x_2$		1	$r_{23}$	$r_{24}$	$r_{25}$	$r_{y2}$
$x_3$			1	$r_{31}$	$r_{32}$	$r_{y3}$
$x_4$				1	$r_{41}$	$r_{y4}$
$x_5$					1	$r_{y5}$
$y$						1

للمثال (٥-١) فإن مصفوفة معاملات الارتباط هي:

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$y$
$x_1$	1	-0.061	0.387	-0.115	0.213	0.339
$x_2$		1	0.106	0.918	-0.111	0.631
$x_3$			1	0.032	0.038	-0.089
$x_4$				1	-0.159	0.413
$x_5$					1	-0.066
$y$						1

من مقارنة معاملات الارتباط البسيطة بين متغير الاستجابة مع كل واحد من المتغيرات المستقلة (العمود الخامس) نجد أن أعلى ارتباط بين  $x_2$ ,  $y$  حيث  $r_{y2} = 0.631$  وعلى ذلك فإن المتغير  $x_2$  هو المتغير المستقل الأول الذي يرشح

للدخول في نموذج الانحدار. معادلة الانحدار المقدرة المحتوية على المتغير  $x_2$  هي:

$$\hat{y} = 2273.088 + 0.08x_2$$

٢- نختبر معنوية المتغير  $x_2$  وذلك باستخدام قيمة  $t$  المحسوبة والخاصة بالمتغير  $x_2$  وهي  $t=3.148$  عند مستوى معنوية  $\alpha = 0.05$  وبما أن  $|t| = 3.148$  تزيد عن قيمة الجدولية حيث  $t_{(15)} = 2.131$  فهذا يعني أن المتغير  $x_2$  له تأثير معنوي على الانحدار لذلك نثبت  $x_2$  في النموذج. معامل التحديد  $R^2_2$  هو 0.398 وهذا يعني أن معادلة الانحدار المقدرة التي حصلنا عليها تقس 39% من الانحرافات الكلية في قيمة  $Y$ .

#### الخطوة الثانية:

١. نحسب مصفوفة معامل الارتباط الجزئي لبقية المتغيرات الغير موجوده في معادلة الانحدار وهي ( $x_1, x_3, x_4, x_5$ ) مع متغير الاستجابة  $Y$  باعتبار ان  $x_2$  ثابت للمثال (٥-١) فهي:

	$x_1$	$x_3$	$x_4$	$x_5$	$y$
$x_1$	1	0.397	-0.150	0.208	0.487
$x_3$		1	-0.166	0.050	-0.202
$x_4$			1	-0.144	-0.541
$x_5$				1	0.006
$y$					1

نختار اكبر معامل ارتباط جزئي وهو للمتغير  $x_4$  حيث

$$|r_{y4.2}| = |-0.541|$$

٢. نحصل على معادلة الانحدار المقدرة المحتوية على المتغيرين  $x_4, x_2$  وهي:

$$\hat{y} = 4600.805 + 0.203x_2 - 21.567x_4.$$

نختبر معنوية المتغير المستقل  $x_4$  وذلك بإستخدام قيمة  $t$  المحسوبة للمتغير  $x_4$  وهي:  $t = -2.408$  وبما أن  $|t| = 2.408$  تزيد عن قيمة  $t$  الجدولي  $t_{0.025}(14) = 2.145$  فهذا يعني أن المتغير  $x_4$  له تأثير معنوي على الانحدار لذلك نثبت  $x_4$  في النموذج. معامل التحديد  $R_3^2 = 0.574$  وهذا يعني أن نموذج الانحدار المقدر في  $x_4$  ،  $x_2$  يفسر 57.4% من الانحرافات الكلية في قيم  $Y$ . أي أن قيمته زادت عن الخطوة السابقة، معنى ذلك أن  $x_4$  قد اضاف جزءاً معنوياً إلى مجموع المربعات المفسرة.

### الخطوة الثالثة:

١٠ يتم حساب مصفوفة معامل الارتباط الجزئي لبقية المتغيرات أي  $(x_1, x_3, x_5)$  مع متغير الاستجابة  $Y$  باعتبار  $x_2, x_4$  ثابتان وهي:

	$x_1$	$x_3$	$x_5$	$y$
$x_1$	1	0.3815	0.1907	0.4888
$x_3$		1	0.0268	-0.3513
$x_5$			1	-0.0864
$y$				1

نختار المتغير الذي له أعلى معامل ارتباط جزئي. لمثالنا فإن أعلى معامل ارتباط جزئي يكون للمتغير  $x_1$  حيث  $r_{y1.24} = 0.4888$ . نحصل على معادلة الانحدار المقدره المحتويه على المتغيرات  $x_1, x_2, x_4$  حيث:

$$\hat{y} = 3653.250 + 10.159x_1 + 0.192x_2 - 19.089x_4.$$

٢٠ نختبر معنوية المتغير  $x_1$  وذلك بإستخدام قيمة  $t$  المحسوبة للمتغير  $x_1$  وهي  $t = 2.02$  وبما أن  $|t| = 2.02$  أقل من القيمة  $t$  الجدولي  $t_{0.025}(13) = 2.16$  وهذا يعني أن  $x_1$  ليس له تأثير معنوي على الانحدار. أي أن مساهمته غير فعالة ويجب حذفه من معادلة الانحدار وبذلك تنتهي خطوات هذه الطريقة وتكون افضل معادلة إنحدار هي:

$$\hat{y} = 4600.805 + 0.203x_2 - 21.567x_4$$

### (٩-٥) طريقة الاختيار التكريري

إن هذه الطريقة التي وضعها (1960) Efroymsen مساهمي التحسين لطريقة الاختيار الامامية وهذا التحسين يتضمن اعادة اختبار مدى أهمية المتغيرات التي تم اختيارها في المراحل السابقة حيث أن المتغير الذي يكون افضل المتغيرات في مرحلة ما قد يكون غير ضروري في مرحلة لاحقة وذلك بسبب الارتباط بينه وبين المتغيرات التي ادخلت حديثا في نموذج الانحدار. ولاختيار ذلك يتم استخدام قيمة  $F$  الجزئية (أو قيمة  $t$ ) لكل متغير في نموذج الانحدار في أي مرحلة على أساس أن هذا المتغير هو أحدث متغير ادخل في النموذج بغض النظر عن نقطة دخوله الحقيقية في النموذج وأي متغير تكون مساهمته غير معنوية يتم حذفه من النموذج وتستمر هذه العملية حتى لايتبقى لدينا أي متغير يقبل في النموذج وكذلك لا يوجد أي متغير يرفض من النموذج. وبتطبيق هذه الطريقة على المثال (٥-١) للوصول الى أفضل نموذج انحدار نتبع الخطوات الآتية:

#### الخطوة الأولى:

١. نبدأ بإيجاد مصفوفة معاملات الارتباط البسيطة بين جميع المتغيرات المستقلة المرشحة ومتغير الاستجابة ويتم اختيار المتغير الذي له أعلى معامل ارتباط مع متغير الاستجابة  $Y$ . للمثال (٥-١) فإن مصفوفة معاملات الارتباط البسيطة تم الحصول عليها عند تقاولنا لطريقة الاختيار الامامية حيث المتغير  $x_2$  هو الذي له أعلى معامل ارتباط مع  $Y$  حيث  $r_{y2}=0.631$ .

٢. يتم تقدير معادلة الانحدار المقدره في المتغير  $x_2$ . للمثال (٥-١) فإن معادلة الانحدار المقدره في المتغير  $x_2$  هي:

$$\hat{y} = 2273.088 + 0.080x_2$$

٣. نختبر معنوية المتغير المستقل  $x_2$  وذلك بحساب قيمة  $t$  الخاصه به وهي  $t=3.148$  وبما أن  $|t|=3.148$  تزيد من القيمة الجدولية لـ  $t$  وهي  $t_{0.025}(15)=2.131$  فهذا يعني أن هذا المتغير معنوي وله مساهمة في نموذج الانحدار وعلى ذلك يتم الاحتفاظ به في نموذج الانحدار.

#### الخطوة الثانية:

١. نوجد مصفوفة الارتباط الجزئية للمتغيرات الباقية مع  $Y$  وهي  $x_3, x_4, x_5$ , بإعتبار أن  $x_2$  ثابت ويتم اختيار المتغير الذي له أعلى معامل ارتباط جزئي. للمثال (٥-١) ومن مصفوفة معاملات الارتباط الجزئية للمتغيرات  $x_3, x_4, x_5$  مع  $Y$  وإعتبار أن  $x_2$  ثابت تم الحصول عليها

عند تناول طريقة الاختيار الامامية حيث المتغير  $x_4$  هو الذي له اعلى ارتباط جزئي مع  $y$  حيث  $|r_{y4.2}| = |-0.54|$  .

٢. نختبر معنوية المتغير المستقل  $x_4$  بإعتبار أنه آخر متغير ادخل في نموذج الانحدار وذلك بحساب قيمة  $t$  لهذا المتغير حيث  $t = -2.408$  وبما أن  $|t| = 2.408$  تزيد عند قيمة  $t$  الجدولية  $t_{0.025}(14) = 2.145$  وعلى ذلك يتم الاحتفاظ بالمتغير  $x_4$  في معادلة الانحدار. ثم نختبر مدى معنوية المتغير المستقل  $x_2$  على أساس أن  $x_4$  قد ادخل اولاً في نموذج الانحدار (وهنا نجد الفرق بين طريقة الانحدار التدرجي وطريقة الاختيار الامامية) وبما أن قيمة  $t$  للمتغير  $x_2$  هي  $t = 3.642$  وبما أن  $|t| = 3.642$  التي تزيد عن قيمة  $t$  الجدولية  $t_{0.025}(14) = 2.145$  فهذا يعني أن المتغير  $x_2$  مازال معنوي وله مساهمة في نموذج الانحدار وعلى ذلك يتم الاحتفاظ به في نموذج الانحدار.

#### الخطوة الثالثة:

١. نوجد مصفوفة الارتباط الجزئية للمتغيرات الباقية مع  $y$  وهي  $x_1, x_3, x_5$  بإعتبار أن  $x_4, x_2$  ثوابت ويتم منها اختيار المتغير الذي له أعلى معامل ارتباط جزئي. للمثال (١-٥) ومن مصفوفة معاملات الارتباط الجزئي التي اوجدناها في طريقة الانحدار الامامية فإن المتغير الذي له أكبر معامل ارتباط جزئي مع  $y$  هو  $x_1$  حيث  $r_{y1.24} = 0.4888$  .

٢. يتم ايجاد معادلة الانحدار المقدره في المتغيرات  $x_1, x_2, x_4$  للمثال (١-٥) فإن:

$$\hat{y} = 3653.250 + 10.159x_1 + 0.192x_2 - 19.089x_4 .$$

٣. نختبر معنوية المتغير المستقل  $x_1$  بإعتباره اخر متغير ادخل في معادلة الانحدار وبما أن قيمة  $t$  المحسوبه للمتغير  $x_1$  هي  $t = 2.02$  وبما أن  $|t| = 2.02$  اقل من قيمة  $t$  الجدولية  $t_{0.025}(13) = 2.16$  فهذا يعني أن هذا المتغير غير معنوي أي أن مساهمته غير فعالة ويجب حذفه من معادلة الانحدار وبذلك تنتهي خطوات هذه الطريقة وتكون أفضل معادلة انحدار مقدره هي:

$$\hat{y} = 4600.805 + 0.203x_2 - 21.567x_4 .$$

حيث قيمة  $t$  لكل من  $x_2, x_4$  معنوية ، وما يجدر الإشارة اليه أنه يجب عدم حذف متغيرين في نفس الوقت .

## الفصل السادس

### نماذج انحدار كثرات الحدود Polynomial Regression Models

- (١-٦) نماذج انحدار كثرات الحدود - متغير مستقل واحد
- (١-١-٦) تقدير المعالم باستخدام المربعات الصغرى
- (٢-١-٦) اختبارات الفروض
- (٣-١-٦) تحديد درجة النموذج
- (٤-١-٦) تحديد القيم المثلى
- (٥-١-٦) انحدار بدلالة الانحرافات
- (٦-١-٦) كثرات الحدود المتعامدة
- (٢-٦) نماذج انحدار كثرات الحدود - متغيرين مستقلين

## (١-٦) نماذج إحداد كثيرات الحدود - متغير مستقل واحد

النماذج التي تناولناها في البند (٢-٤) تشتمل على نوال في متغير مستقل  $x$  حيث تكون الدالة متزايدة باضطراب أو متناقصة باضطراب. في حالات كثيرة ولأسباب نظرية أو بسبب شكل الانتشار للبيانات يقترح أن نموذج الإحداد الحقيقي  $\mu_{y|x}$  له واحد أو أكثر من قيمة عظمى أو صغرى حيث:

$$\mu_{y|x} = \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k. \quad (١-٦)$$

الآن المطلوب إيجاد معادلة الإحداد المقدره لـ  $\mu_{y|x}$  في (١-٦) بالاعتماد على أزواج المشاهدات التي عددها  $n$  حيث  $\{(x_j, y_j), j=1, 2, \dots, n\}$ . كل مشاهد  $y_j$  تحقق للمعادلة:

$$y_j = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \dots + \beta_k x_j^k + e_j^*.$$

أو:

$$y_j = b_0 + b_1 x_j + b_2 x_j^2 + \dots + b_k x_j^k + e_j.$$

حيث  $k$  تمثل درجة كثيرات الحدود و  $e_j^*, e_j$  مره أخرى الخطأ العشوائي والباقي المرتبطين بالاستجابة  $y_j$ . هنا عدد الأزواج ( $n$ ) لابد أن يكون على الأقل أكبر من  $p = k+1$  حيث  $p$  تمثل عدد المعالم المقدره.

نموذج الإحداد في هذه الحالة يأخذ الشكل التالي:

$$Y_j = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \dots + \beta_k x_j^k + \epsilon_j, j=1, 2, \dots, n. \quad (٢-٦)$$

نموذج الإحداد المقدر للنموذج (٢-٦) يأخذ الشكل التالي:

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + \dots + b_k x^k. \quad (٣-٦)$$

إذا كانت  $k=1$  في (٢-٦) فإن النموذج يصبح نموذج انحدر خطي بسيط. أما إذا كانت  $k=2$  في (٢-٦) فإن النموذج في هذه الحالة يسمى نموذج إحدار من الدرجة الثانية (أو النموذج التربيعي) وهو من أبسط أنواع نماذج الإحداد متعدد الحدود حيث يضم النموذج المتغير  $x$  و المتغير  $x^2$ . يستخدم نموذج الإحداد من الدرجة الثانية في الحالات التالية:



• عندما تكون دالة المتغير التابع الحقيقية دالة من الدرجة الثانية تضم مكوّن الأثر الخطي والتربيعي معاً.

• عندما تكون دالة الانحدار مجهولة أو معقدة والنموذج من الدرجة الثانية يمثل أفضل تقدير للدالة المجهولة.

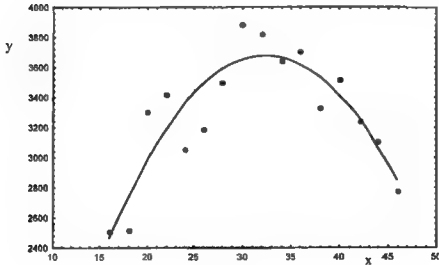
عندما  $k=3$  في النموذج (٢-٦) فيسمى النموذج بنموذج الانحدار من الدرجة الثالثة أو (النموذج التكعبي). إن شكل معادلة الانحدار المقدره يتأثر بدرجة نموذج الانحدار فإذا كانت معادلة الانحدار المقدره في (٣-٦) من الدرجة الثانية (التربيعية) أي:

$$\hat{y} = b_0 + b_1x + b_2x^2.$$

فإن المنحنى الناتج ينحني مره واحده فقط أما للأعلى كما في شكل (١-٦) أو للأسفل كما في شكل (٢-٦). إذا كان منحنى معادلة الانحدار المقدره (٣-٦) من الدرجة الثالثة (التكعبيية) فيصبح:

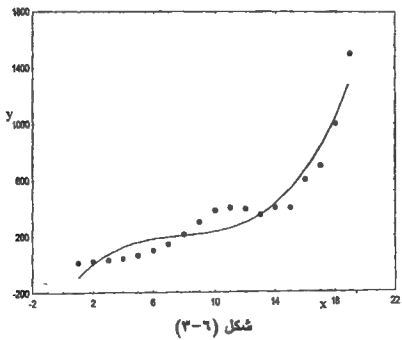
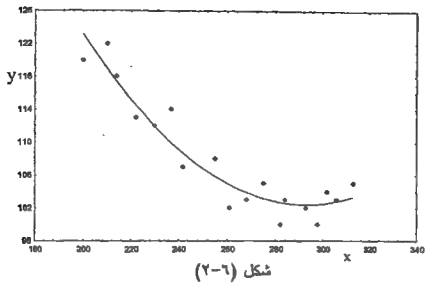
$$\hat{y} = b_0x + b_1x + b_2x^2 + b_3x^3$$

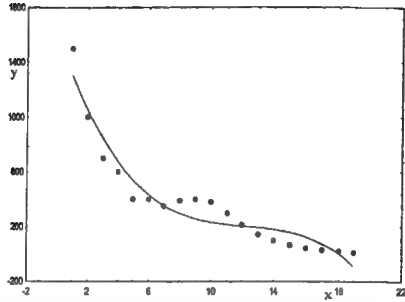
فإنه ينحني مره للأعلى وأخرى للأسفل كما في شكل (٣-٦) أو العكس كما في شكل (٤-٦).



شكل (١-٦)

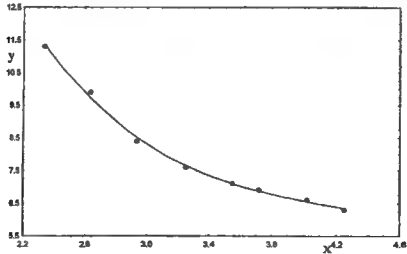
- ۳۷۱ -





شكل (٦-٤)

وبصوره عامه فإن منحنى المعادلة من الدرجة  $k$  ينحني  $(k-1)$  من المرات مع ملاحظة أنه ليس من الضروري أن ينحني بالضبط  $(k-1)$ . فمثلاً المنحنى في شكل (٦-٥) يمثل معادلة إتحاداً مقدره من الدرجة الثالثة وفيه انحناء واحد بسيط فقط.



شكل (٦-٥)

(١-٦-١) تقدير المعالم باستخدام المربعات الصغرى

لتقدير المعالم  $\beta_0, \beta_1, \dots, \beta_k$  في (١-٦) فإننا نوجد التقديرات  $b_0, b_1, \dots, b_k$  التي تؤدي إلى تصغير مجموع مربعات البواقي التالية:

$$\sum_{j=1}^n [y_j - (b_0 + b_1 x_j + b_2 x_j^2 + \dots + b_k x_j^k)]^2 .$$

والتي نحصل عليها بحل المعادلات الطبيعية التالية آنيا:

$$b_0 n + b_1 \sum x_j + b_2 \sum x_j^2 + \dots + b_k \sum x_j^k = \sum y_j$$

$$b_0 \sum x_j + b_1 \sum x_j^2 + b_2 \sum x_j^3 + \dots + b_k \sum x_j^{k+1} = \sum x_j y_j$$

$$\vdots \quad \vdots \quad \vdots$$

$$b_0 \sum x_j^k + b_1 \sum x_j^{k+1} + b_2 \sum x_j^{k+2} + \dots + b_k \sum x_j^{2k} = \sum x_j^k y_j .$$

مثال (١-٦)

لأزواج المشاهدات المعطاة في جدول (١-٦) لوجد معادلة الانحدار المقدره لنموذج انحدار من الدرجة الثانية.

جدول (١-٦)

x	1	2	3	4	5	6	7	8	9	10
y	5.0	10.20	15.35	20.50	25.95	32.20	38.50	46.00	53.80	62.00

الحل

من البيانات في جدول (١-٦) فإن:

$$n = 10 , \sum x_i = 55 , \sum y_i = 309.5,$$

$$\sum x_i y_i = 2218.1 , \sum x_i^2 = 385,$$

$$\sum x_i^2 y_i = 17708.2 , \sum x_i^3 = 3025 , \bar{y} = 30.95,$$

$$\sum x_i^4 = 25333 , \sum y_i^2 = 12831.845,$$

تستخدم القيم السابقة في الحصول على المعادلات التالية:

$$10b_0 + 55b_1 + 385b_2 = 309.5$$

$$55b_0 + 385b_1 + 3025b_2 = 2218.1$$

$$385b_0 + 3025b_1 + 25333b_2 = 17708.2$$

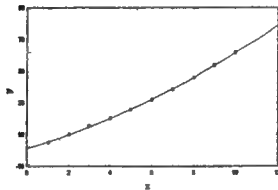
بحل المعادلات السابقة آنيا يمكن إيجاد  $b_0$ ,  $b_1$ ,  $b_2$ . حل لهذه المعادلات هو:

$$b_0 = 1.48083, b_1 = 3.792313, b_2 = 0.223674$$

معادلة الانحدار المقدرة هي:

$$\hat{y} = 1.48083 + 3.792313x + 0.223674x^2$$

والمصنعة بيئنا مع شكل الانتشار في شكل (٦-٦).



شكل (٦-٦)

هذا ويمكن الحصول على التقديرات  $b_0, b_1, \dots, b_k$  باستخدام أسلوب المصفوفات وذلك كما في الانحدار الخطي المتعدد كما ذكرنا سابقا حيث :

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^k \\ 1 & x_2 & x_2^2 & \dots & x_2^k \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^k \end{bmatrix}$$

حيث  $x_j$  تمثل المشاهدات رقم  $j = 1, 2, \dots, n$  وعليه فإن:

$$X'X = \begin{bmatrix} n & \sum x_j & \sum x_j^2 & \dots & \sum x_j^k \\ \sum x_j & \sum x_j^2 & \sum x_j^3 & \dots & \sum x_j^{k+1} \\ \sum x_j^2 & \sum x_j^3 & \sum x_j^4 & \dots & \sum x_j^{k+2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \sum x_j^k & \sum x_j^{k+1} & \sum x_j^{k+2} & \dots & \sum x_j^{2k} \end{bmatrix}$$

وعلى ذلك تقدير المربعات الصغرى لـ  $b$  هو

$$b = (X'X)^{-1} X'y.$$

ولتسهيل العمليات الحسابية فإنه يمكن استبدال  $x_j^k$  في نموذج الانحدار (٢-٦) بالمتغير  $x_{kj}$  بحيث يصبح (٢-٦) نموذج انحدار خطي متعدد لـ  $k$  من المتغيرات حيث  $x_{1j} = x_j, x_{2j} = x_j^2, \dots, x_{kj} = x_j^k$  وعلى ذلك تصبح المصفوفة  $X'X$  كالآتي:

$$X'X = \begin{bmatrix} n & \sum x_j & \sum x_{2j} & \dots & \sum x_{kj} \\ \sum x_{1j} & \sum x_{1j}^2 & \sum x_{1j}x_{2j} & \dots & \sum x_{1j}x_{kj} \\ \sum x_{2j} & \sum x_{2j}x_{1j} & \sum x_{2j}^2 & \dots & \sum x_{2j}x_{kj} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \sum x_{kj} & \sum x_{kj}x_{1j} & \sum x_{kj}x_{2j} & \dots & \sum x_{kj}^2 \end{bmatrix}$$

- ١٥١ -

$$X' y = \begin{bmatrix} \sum y_j \\ \sum x_{1j} y_j \\ \sum x_{2j} y_j \\ \vdots \\ \sum x_{kj} y_j \end{bmatrix},$$

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix},$$

$$b = (X'X)^{-1} X' y$$

يمكن استخدام برامج الحاسب الآلي الجاهزة والخاصة بالانحدار في إيجاد التقديرات  $b_0, b_1, \dots, b_k$ .

### ٦-١-٢) اختبارات الفروض

١- اختبار معنوية الانحدار ككل:

إن هذا الاختبار يستخدم لاختبار فرض العدم:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

ضد الفرض البديل على الأقل واحد من المعاملات الجزئية لا يساوي صفر. سوف نستخدم الإحصاء  $F$  على الصورة التالية:

$$F = \frac{MSR(\beta_1, \beta_2, \dots, \beta_k | \beta_0)}{MSE(\beta_1, \beta_2, \dots, \beta_k | \beta_0)}$$

حيث:

$MSR$ : متوسط مجموع مربعات الانحدار لنموذج الانحدار من الدرجة  $k$ .

$MSE$ : متوسط مجموع مربعات البواقي لنموذج الانحدار من الدرجة  $k$ .

$n$  : عدد المشاهدات.

$k$  : عدد المتغيرات المستقلة  $(x_1, x_2)$ .

ويمكن الوصول لقرار بشأن معنوية الانحدار ككل كمايلي:

- إذا كانت قيمة  $F$  المحسوبة تزيد عن قيمة  $F$  الجدوليه بدرجة حرة  $k, (n-k-1)$  ومستوى معنوية  $\alpha$  نرفض فرض العدم ونقبل الفرض البديل ومن ثم فإن كل قيم معاملات النموذج لا تساوي الصفر. أي أن المتغيرين  $(x_1, x_2)$  لهما تأثير معنوي على المتغير التابع.
  - إذا كانت قيمة  $F$  المحسوبة أقل من القيمة الجدولية نقبل فرض العدم بتساوي كل معاملات النموذج للصفر أي أن الانحدار غير معنوي بمعنى أن المتغيرين لا يؤثران على متغير الاستجابة.
- وعادة تنظم النتائج في جدول تحليل التباين كما هو موضح في جدول (٦-٢).

جدول (٦-٢)

S.O.V	df	SS	MS	F
الانحدار	k	$SSR(\beta_1, \beta_2, \dots, \beta_k   \beta_0) = SSR$	MSR	$F = \frac{MSR}{MSE}$
الخطأ	n-k-1	SSE	MSE	
الكل	n-1	SYT		

#### مثال (٦-٢)

لأزواج المشاهدات المعطاة في جدول (٦-٣) حيث  $x$  تمثل عدد الأيام بعد الإزهار و  $y$  (Kg/ha) المحصول الناتج من نبات ما في الهند. أوجد معادلة الانحدار المقدره من الدرجة الثانية واختبر فرض العدم:

$$H_0 : \beta_1 = \beta_2 = 0.$$

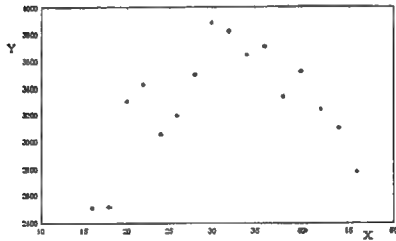


جدول (٦-٣)

x	16	18	20	22	24	26	28	30
y	2508	2518	3304	3423	3057	3190	3500	3883
x	32	34	36	38	40	42	44	46
y	3823	3646	3708	3333	3517	3241	3103	2776

الحل

شكل الانتشار موضح في شكل (٦-٧).

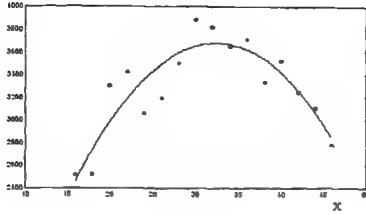


شكل (٦-٧)

والذي يوضح علاقة من الدرجة الثانية. نموذج الإحداد المقدر باستخدام برنامج

SPSS هو:  $\hat{y} = -1070.4 + 293.48x - 4.5358x^2$ .

والموضح بيانيا في شكل (٦-٨) مع شكل الانتشار.



شكل (٦-٨)

جدول تحليل التباين معطى في جدول (٦-٤)  
جدول (٦-٤)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2   \beta_0$	2	2084779.4	1042389.691	25.077
الخطأ	13	540388.37	41568.336	
الكلي	15	2625167.8		

بما أن قيمة F المحسوبة من جدول (٦-٤) تساوي (25.077) تزيد عن قيمة F الجدولية عند درجتي حرية (2, 13) و  $(F_{0.05}(2,13) = 3.81, \alpha = 0.05)$  فإننا نرفض فرض العدم.

٢- اختبار حول معامل التقدير جزئي معين:

لاختبار فرض العدم:

$$H_0 : \beta_1 = 0.$$

علينا افترض أن:

$$\beta_{i+1} = \beta_{i+2} = \dots = \beta_k = 0.$$

فإننا نستخدم واحد من الاختبارين التاليين:

• اختبار  $F$  الجزئي حيث الإحصاء  $F$  يأخذ الصيغة التالية:

$$F = \frac{MSR(\beta_i | \beta_0, \beta_1, \beta_2, \dots, \beta_{i-1})}{MSE(\beta_1, \beta_2, \dots, \beta_i)}.$$

أي متجهين  $x^k, x^{i+2}, \dots, x^{i+1}$ . نرفض فرض العدم إذا كانت قيمة  $F$  المحسوبة تزيد عن القيمة الجدولية المعينة.

مثال (٦-٣)

للمثال (٦-٢) لاختبار فرض العدم:

$$H_0: \beta_2 = 0.$$

وذلك تحت الفرض أن نموذج الانحدار من الدرجة الثالثة نفع الآتي:

عندما يكون الحد  $x$  فقط في نموذج الانحدار المقدر فإن جدول تحليل التباين معطى في جدول (٦-٥).

جدول (٦-٥)

S.O.V	df	SS	MS	F
$\beta_1   \beta_0$	1	204526.24	204526.238	1.183
الخطأ	14	2420641.5	172902.965	
الكلية	15	2625167.8		

عند وجود الحدين  $x^2, x$  في النموذج فإن جدول تحليل التباين معطى في جدول (٦-٦).

جدول (٦-٦)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2   \beta_0$	2	2084779.4	1042389.691	25.077
الخطأ	13	540388.37	41568.336	
الكلية	15	2625167.8		

من جدول (٥-٦) وجدول (٦-٦) نحصل على  $SSR(\beta_2 | \beta_1, \beta_0)$  كالتالي:

$$SSR(\beta_2 | \beta_1, \beta_0) = SSR(\beta_1, \beta_2 | \beta_0) - SSR(\beta_1 | \beta_0)$$

$$= 2084779.4 - 204526.24$$

$$= 1880253.16.$$

$$MSR(\beta_2 | \beta_1, \beta_0) = SSR(\beta_1, \beta_2 | \beta_0) / 1.$$

$$= 1880253.16,$$

حيث الواحد هو درجة الحرية المرتبطة بـ  $MSR(\beta_2 | \beta_1, \beta_0)$ .  
وعلى ذلك قيمة  $F$  هي:

$$F = \frac{MSR(\beta_2 | \beta_1, \beta_0)}{MSE(\beta_1, \beta_2 | \beta_0)}$$

$$= \frac{1880253.16}{41568.336} = 45.2333,$$

حيث  $MSE(\beta_1, \beta_2 | \beta_0)$  مأخوذة من جدول (٦-٦). وبما أن قيمة  $F$  المحسوبة تزيد عن قيمة  $F$  الجدولية  $F_{0.05}(1, 13) = 4.67$  فإننا نرفض فرض العدم.

• اختبار  $t$  حيث :-

$$t = \frac{b_j}{s.e(B_j)}$$

تحت فرض تجاهل:

$$\beta_{i+1}, \beta_{i+2}, \dots, \beta_k.$$

حيث  $s.e(B_i)$  هو تقدير للانحراف المعياري للمقدر  $B_i$ . وأن:

$$s.e(B_i) = \sqrt{MSE c_{ii}}.$$

حيث MSE مستخرجه من جدول تحليل التباين في جدول (٦-٦) عند وجود الحدود  $x^2, \dots, x^k$  في النموذج و  $c_{ii}$  هو آخر حد قطري في المصفوفة  $(X'X)^{-1}$ .

• عند استخدام اختبار  $t$  فإن قيمة  $t$  المحسوبة سوف تكون:

$$t = \frac{b_2}{s.e(B_2)} = \frac{-4.536}{0.674} = -6.726.$$

وبما أن  $|t|$  المحسوبة تزيد عن قيمة  $t$  الجدوليّه عند درجات حرية 13 و  $\alpha = .05$  حيث  $t_{.025}(13) = 2.16$  فإننا نرفض فرض العدم ويجب أن نعلم أن  $F = t^2$  الجزئية حيث  $t^2 = 45.239$  وهي نفسها التي حصلنا عليها من اختبار  $F$  الجزئية.

يفيد هذا الاختبار في تحديد درجة المعادلة كما سيأتي ذكره بعد قليل. كما أن هذا الاختبار يختلف عن الاختبار المماثل له في موضوع الاتحاد الخطي المتعدد. فكما ذكرنا من قبل قليل فإنه لا اختبار  $H_0: \beta_i = 0$  في الاتحاد لكثيرات الحدود فإننا نفترض أن:

$$\beta_{i+1} = \beta_{i+2} = \dots = \beta_k = 0.$$

ونستخدم الإحصاء  $F$  على الصورة:

$$F = \frac{MSR(\beta_i | \beta_1, \beta_2, \dots, \beta_{i-1}, \beta_0)}{MSE(\beta_1, \beta_2, \dots, \beta_i | \beta_0)}$$

أما في الاتحاد الخطي المتعدد فإننا لا نهمل:

$$\beta_{i+1} = \beta_{i+2} = \dots = \beta_k$$

ونستخدم الإحصاء  $F$  على الصورة:

$$F = \frac{MSR(\beta_i | \beta_1, \beta_2, \dots, \beta_{i-1}, \beta_{i+1}, \dots, \beta_k, \beta_0)}{MSE(\beta_1, \beta_2, \dots, \beta_k | \beta_0)}.$$

كما أن الهدف من اختبار  $\beta_j = 0$  في انحدار كثيرات الحدود هو لتحديد درجة المعادلة، بينما هذا الاختبار (أي  $\beta_j = 0$ ) في الانحدار الخطي المتعدد يعني ما إذا كانت إضافة  $x_j$  في النموذج ستساعد معنوياً في التنبؤ بمتوسط  $Y$  بوجود غيره من المتغيرات وهي  $x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_k$ .

### ٣- اختبار فروض أخرى

بفرض أن لدينا نموذج انحدار من الدرجة الثالثة:

$$Y_j = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \beta_3 x_j^3 + \epsilon_j$$

لاختبار فرض العدم:

$$Y_j = \beta_2 x_j^2 + \epsilon_j$$

فإننا نستخدم الإحصاء  $F$  على الصورة :

$$F = \frac{MSR(\beta_1, \beta_3 | \beta_2, \beta_0)}{MSE(\beta_1, \beta_2, \beta_3 | \beta_0)}$$

للمثال (٦-٢) وبفرض أن لدينا نموذج انحدار من الدرجة الثالثة ويرغب الباحث في اختبار فرض العدم:

$$H_0 : Y = \beta_2 x^2 + \epsilon$$

جدول تحليل للتباين عندما  $x, x^2, x^3$  في نموذج الانحدار معطى في جدول (٦-٧)  
جدول (٦-٧)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2, \beta_3   \beta_0$	3	2091716.8	697238.949	15.684
الخطأ	12	533450.90	44454.242	
الكلي	15	2625167.8		

جدول تحليل التباين عندما  $x^2$  فقط في النموذج معطى في جدول (٦-٨).

جدول (٦-٨)

S.O.V	df	SS	MS	F
$\beta_2 \beta_0$	1	72149.560	72149.560	0.396
الخطأ	14	2553018.2	182358.442	
الكل	15	2625167.8		

من جدول (٦-٧) و جدول (٦-٨) فإن:

$$SSR(\beta_1, \beta_3|\beta_2, \beta_0) = SSR(\beta_1, \beta_2, \beta_3|\beta_0) - SSR(\beta_2|\beta_0)$$

$$= 2091716.8 - 72149.560$$

$$= 2019567.24.$$

$$MSR(\beta_1, \beta_3|\beta_2, \beta_0) = 2019567.24 / 2$$

$$= 1009783.62 .$$

وعلى ذلك قيمة F المحسوبة هي:

$$\begin{aligned} F &= \frac{MSR(\beta_1, \beta_3|\beta_2, \beta_0)}{MSE(\beta_1, \beta_2, \beta_3|\beta_0)} \\ &= \frac{1009783.62 / 2}{44454.242} \\ &= 11.3576. \end{aligned}$$

حيث  $MSE(\beta_1, \beta_2, \beta_3|\beta_0)$  معطاة في جدول (٦-٧). وبما أن قيمة F المحسوبة تزيد عن قيمة F الجدولية  $F_{0.05} [2, 12] = 3.89$  فإننا نرفض فرض العدم.

## ٢-١-٦) تحديد درجة النموذج

سوف نقدم ثلاثة طرق لتحديد درجة النموذج:

### ١. اختبار نقص التوفيق

ويستخدم هذا الاختبار عندما يكون هناك مشاهدات متكررة على قيم الاستجابة  $Y$  وذلك على الأقل لمستوى واحد من  $x$ . بفرض أننا أخذنا عينة عشوائية من  $n$  من المشاهدات وذلك باستخدام  $m$  من القيم المختلفة  $x$  بحيث أن العينة تحتوي على  $n_1$  قيمة مشاهدة من المتغير العشوائي  $Y_1$  المقابل لـ  $x_1$  و  $n_2$  قيمة مشاهدة من المتغير العشوائي  $Y_2$  المقابل لـ  $x_2$  و .... و  $n_m$  قيمة مشاهدة من المتغير العشوائي  $Y_m$  المقابل لـ  $x_m$ . من الضروري أن  $n = \sum n_i$ . وتتلخص هذه الطريقة فيما يلي:

• نبدأ بافتراض أن نموذج الانحدار هو:

$$Y = \beta_0 + \beta_1 x + \epsilon.$$

• نختبر فرض العدم

$$H_0: \beta_1 = 0.$$

• نجزأ مجموع مربعات البواقي إلى قسمين: الأول مجموع المربعات الذي يعود إلى نقص التوفيق والآخر يعود إلى الخطأ الخالص كما أوضحنا في الفصل الثاني عند اختبار خطية النموذج. يستخدم في اختبار الاحصاء  $F$  على الصورة:

$$F = \frac{MSLF}{MSPE}.$$

حيث  $MSLF$  متوسط مجموع المربعات الذي يعود إلى نقص التوفيق و  $MSPE$  متوسط مجموع المربعات الذي يعود إلى الخطأ الخالص. إذا كانت قيمة  $F$  المحسوبة تزيد عن القيمة الجدولية المعينة. يضاف الحد  $x^2$  إلى نموذج الانحدار البسيط ليصبح النموذج من الدرجة الثانية (نموذج تربيعي) حيث:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon.$$



- نجزاً مجموع مربعات البواقي إلى قسمين كما في حالة نموذج الانحدار البسيط وإذا كانت قيمة  $F$  المحسوبة تزيد عن القيمة الجدولية نضيف الحد  $x^3$  وهكذا نستمر حتى نحصل على  $F$  غير معنوية فننتوقف عن إضافة حدود أخرى.

مثال (٦-٤)

درست فعالية (جير) تجريبي جديد في تخفيض الجازولين في 12 محاولة استخدمت فيها عربة نقل خفيفة مجهزة بهذا الجير حيث  $x$  في جدول (٦-٩) للسرعة الثابتة (بالميل في الساعة) لعربة الاختبار و  $y$  الأميال المقطوعة لكل جالون.

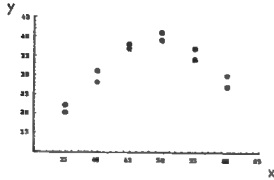
جدول (٦-٩)

$x$	$y$
35	22
35	20
40	28
40	31
45	37
45	38
50	41
50	39
55	34
55	37
60	27
60	30

المطلوب: تحديد درجة نموذج الانحدار باستخدام اختبار نقص التوافق.

**الحل**

شكل الانتشار للبيانات في جدول (٩-٦) معطاة في شكل (٩-٦).



شكل (٩-٦)

تحت فرض نموذج الانحدار البسيط فإن معادلة الانحدار المقترنة هي:

$$\hat{y} = 16.2571 + 0.331429x.$$

جدول تحليل التباين معطى في جدول (١٠-٦)

جدول (١٠-٦)

S.O.V	df	SS	MS	F
الانحدار	1	96.1143	96.114	2.32224
الخطأ	10	413.886	41.3886	-
الكلية	11	510	-	-

الآن لاختبار نقص للتوفيق أي اختبار فرض العدم:

$$H_0 : \mu_{Y|x} = \beta_0 + \beta_1 x.$$

ضد الفرض البديل :

$$H_1 : \mu_{Y|x} \neq \beta_0 + \beta_1 x.$$

نحسب الآتي:

مجموع مربعات الخطأ الخالص عند  $x = 35$  هو:

$$(22)^2 + (20)^2 - \{(22+20)^2/2\} \\ = 2.$$

بدرجات حرية  $(n_1 = 2-1=1)$

مجموع مربعات الخطأ الخالص عند  $x = 40$  هو:

$$(28)^2 + (31)^2 - \{(28+31)^2/2\} \\ = 4.5.$$

بدرجات حرية  $(n_2 = 2-1=1)$

بنفس الطريقة يمكن حساب مربعات الخطأ الخالص للقيم الباقية من  $x$  كما هو موضح في جدول (٦-١١).

جدول (٦-١١).

x	$\sum_{j=1}^{n_j} (y_{ij} - \bar{y}_i)^2$	درجات الحرية
35	2	1
40	4.5	1
45	0.5	1
50	2	1
55	4.5	1
60	4.5	1

جدول تحليل التباين معطى في جدول (٦-١٢).

جدول (١٢-٦)

S.O.V	df	SS	MS	F
الانحدار	1	96.1143	96.1134	2.32224
الخطأ	10	413.886	41.3886	-
تقصير التوفيق	4	395.886	98.9714	32.9905
الخطأ الخالص	6	18	3	-

من جدول (١٢-٦) وبما أن قيمة  $F$  المحسوبة لتقصير التوفيق (32.9905) تزيد القيمة الجدولية  $F_{0.05} [4, 6] = 4.53$  فإننا نرفض فرض العدم:  
الآن نختبر فرض العدم :

$$H_0 : \mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2.$$

ضد الفرض البديل :

$$H_1 : \mu_{Y|x} \neq \beta_0 + \beta_1 x + \beta_2 x^2.$$

جدول تحليل التباين معطى في جدول (١٣-٦).

جدول (١٣-٦)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2   \beta_0$	2	483.168	241.583	
$\beta_1   \beta_0$	1	96.1143	96.1143	
$\beta_2   \beta_1 \beta_0$	1	387.054	387.054	
الخطأ	9	26.832	2.981	
تقصير التوفيق	3	8.832	2.944	<1
الخطأ الخالص	6	18	3	

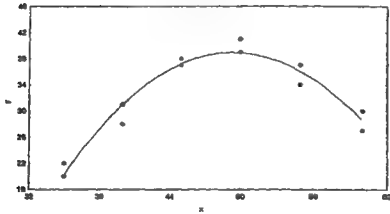
من جدول (٦-١٣) وبما أن قيمة F المحسوبة لقصور التوفيق أقل من الواحد الصحيح فهذا يعني قبول فرض العدم وهو:

$$H_0 : \mu_{y|x} = \beta_0 + \beta_1 x + \beta_2 x^2 .$$

معادلة الانحدار المقدرة هي:

$$\hat{y} = -182.582 + 8.983x - .09107x^2 .$$

والممثلة بيانيا في شكل (٦-١٠) مع شكل الانتشار.



شكل (٦-١٠)

## ٢- الطريقة الانمائية:

ونستخدم عندما لا يكون هناك تكرار لقيم y ونتلخص فيما يلي:

- نفترض نموذج الانحدار البسيط

$$Y = \beta_0 + \beta_1 x + \epsilon .$$

- نختبر فرض العدم

$$H_0 : \beta_1 = 0 .$$

ونلك باستخدام الإحصاء  $F$  حيث:

$$F = \frac{MSR(\beta_1|\beta_0)}{MSE(\beta_1|\beta_0)}.$$

إذا كانت قيمة  $F$  المحسوبة تزيد عن قيمة  $F$  الجدولية المعينة عند مستوى معنوية  $\alpha$  فإننا نضيف الحد  $x^2$  إلى نموذج الإنحدار البسيط ليصبح النموذج من الدرجة الثانية:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon.$$

لاختبار فرض العدم:

$$H_0 : \beta_2 = 0.$$

نستخدم الإحصاء  $F$  حيث:

$$F = \frac{MSR(\beta_2|\beta_1, \beta_0)}{MSE(\beta_1, \beta_2|\beta_0)}.$$

إذا كانت قيمة  $F$  المحسوبة تزيد عن قيمة  $F$  الجدولية المعينة عند مستوى معنوية  $\alpha$  فإننا نضيف الحد  $x^3$  إلى النموذج من الدرجة الثانية. وهكذا إلى أن نحصل على  $F$  غير معنوية لمرتين متتاليتين.

### ٣- الطريقة الخلفية

نتلخص خطواتها فيمايلي:

١- نبدأ بنموذج إنحدار يحتوي على حدود عليا أعلى مايمكن. إن أعلى حد يمكن نظريا البدء به هو درجة  $(k-1)$  حيث  $k$  هي عدد قيم  $x$  الغير متكررة. فمثلا إذا كان عدد قيم  $x$  الغير متكررة هو 10 فإن أعلى درجة للمعادلة تكون 9 و في التطبيق العملي لايمكن حساب معادلة أكثر من الدرجة الرابعة ونلك لصعوبة تفسير معادلات من درجات أعلى. بفرض على سبيل المثال أننا فرضنا معادلة من الدرجة الثالثة. أي :

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon.$$

لاختبار فرض العدم

$$H_0: \beta_3 = 0.$$

نحسب الإحصاء F على الصورة:

$$F = \frac{SSR(\beta_3|\beta_1, \beta_2, \beta_0)/1}{MSE(\beta_1, \beta_2, \beta_3|\beta_0)}.$$

حيث تحسب  $SSR(\beta_3|\beta_1, \beta_2, \beta_0)$  كالآتي:

$$SSR(\beta_3|\beta_1, \beta_2, \beta_0) = SSR(\beta_1, \beta_2, \beta_3|\beta_0) - SSR(\beta_1, \beta_2, \beta_0).$$

إذا كانت قيمة F المحسوبة أقل من F الجدولي المعينة عند مستوى معنوية  $\alpha$  فإن الحد  $\beta_3 x^3$  يحذف من النموذج لتصبح المعادلة من الدرجة الثانية ثم نختبر فرض العدم:

$$H_0: \beta_2 = 0.$$

باستخدام الإحصاء F على الشكل:

$$F = \frac{SSR(\beta_2|\beta_1, \beta_0)/1}{MSE(\beta_1, \beta_2|\beta_0)}.$$

إذا كانت قيمة F المحسوبة أقل من القيمة الجدولية المعينة عند مستوى معنوية  $\alpha$  نحذف الحد  $\beta_2 x^2$  من النموذج وهكذا نستمر بخفض المعادلة بالتدرج حتى نصل إلى الدرجة المناسبة عندما تزيد قيمة F المحسوبة عن القيمة الجدولية

عندئذ نقف عن الحذف ونقرر بأن هذه المعادلة هي التي توافق البيانات.

مثال (٦-٥)

للمثال (٦-٥) المطلوب تحديد درجة نموذج الانحدار.

الحل

من البيانات في جدول (٦-٥) وللخاصة بالمثال (٦-٥) فإن معادلة الانحدار المقدره بفرض نموذج انحدار من الدرجة الثالثة هي:

$$\hat{y} = -203.609 + 199.077x - 1.321x^2 + 0.03457x^3$$

جدول تحليل التباين معطى في جدول (٦-١٤).

جدول (٦-١٤)

S.O.V	df	SS	MS	F
$R(\beta_1, \beta_2, \beta_3   \beta_0)$	3	2091716.8	697238.949	15.684
الخطأ	12	533450.90	44454.242	
الكلية	15	2625167.8		

جدول تحليل التباين عندما  $x, x^2$  فقط في النموذج معطى في جدول (٦-٦).  
وعلى ذلك نحسب :

$$\begin{aligned}
 SSR(\beta_3 | \beta_1, \beta_2, \beta_0) &= SSR(\beta_1, \beta_2, \beta_3 | \beta_0) \\
 &- SSR(\beta_1, \beta_2 | \beta_0) \\
 &= 2091716.8 - 2084779.4 \\
 &= 6937.4.
 \end{aligned}$$

قيمة  $F$  المحسوبة هي:

$$F = \frac{SSR(\beta_3 | \beta_1, \beta_2, \beta_0) / 1}{MSE(\beta_1, \beta_2, \beta_3 | \beta_0)} = \frac{6937.4}{44454.242} = 0.1561.$$

وبما أن قيمة  $F$  الجزئية أقل من الواحد الصحيح نحذف الحد  $\beta_3 x^3$  من المثال (٦-٧) أثبتنا أن قيمة  $F$  حيث:

$$F = \frac{SSR(\beta_2 | \beta_1, \beta_0) / 1}{MSE(\beta_1, \beta_2 | \beta_0)} = 45.2333.$$



معنوية وعلى ذلك نقف عن الحذف ونقرر أن المعادلة من الدرجة الثانية هي التي توافق البيانات. مما يجدر الإشارة إلى أن الطريقة الخلفية يمكن أن تعطي نتائج تختلف عن الطريقة الامامية.

#### (٦-١-٤) تحديد القيم المثلى

من الاستخدامات الأساسية لنموذج الانحدار من الدرجة الثانية هو تحديد القيم المثلى للمتغير التابع والمستقل. ولتحقيق قيمة المتغير المستقل  $x$  التي تحقق أعلى (أدنى) قيمة للمتغير التابع يتم إجراء تفاضل لنموذج الانحدار من الدرجة الثانية بالنسبة لـ  $x$  ومساواة الناتج بالصفر كما يلي:

$$\hat{y} = b_0 + b_1 x + b_2 x^2.$$

$$\frac{\partial \hat{y}}{\partial x} = b_1 + 2b_2 x = 0.$$

إن:

$$x = x_m = \frac{-b_1}{2b_2}.$$

حيث  $x_m$  القيمة المثلى. وللحصول على القيمة العليا (الدنيا) للمتغير التابع يتم التعويض عن قيمة  $x$  في معادلة نموذج الانحدار من الدرجة الثانية بالقيمة المثلى  $x_m$  كما يلي:

$$x_m = \frac{-b_1}{2b_2}.$$

وعلى ذلك:

$$\hat{y}_m = b_0 + b_1 \frac{-b_1}{2b_2} + b_2 \left( \frac{-b_1}{2b_2} \right)^2 = b_0 - \frac{b_1^2}{4b_2}.$$

إن  $\hat{y}_m$  تعد نهاية عظمى إذا كانت إشارة  $b_2$  سالبة ونهاية صغرى إذا كانت إشارة  $b_2$  موجبة. للمثال (٦-٢) حيث:

$$\hat{y} = -1070.4 + 293.48x - 4.5358x^2.$$

فإن:

$$\hat{y}_m = b_0 - \frac{b_1^2}{4b_2} = (-1070.4) - \frac{(293.48)^2}{4(-4.5358)} = 3676.86.$$

وهي نهاية عظمى لأن إشارة  $b_2$  سالبة.

(٦-١-٥) الإحذار بدلالة الانحرافات

بفرض النموذج:

$$Y' = \beta'_0 + \beta'_1 z + \beta'_2 z^2 + \dots + \beta'_k z^k + \varepsilon.$$

حيث  $z = x - \bar{x}$  ,  $Y' = (Y - \bar{Y})$  , لاحظ التعبير عن المتغير المستقل و المتغير التابع كإنحرافات عن  $\bar{x}$  و  $\bar{Y}$  على التوالي وسبب استخدام الانحرافات حول المتوسط في نماذج انحذار كثيرات الحدود هو أن  $x^2$  ,  $x^3$  والحدود من قوى أعلى ستكون في الغالب مرتبطة ارتباطا عاليا والتي تسبب صعوبات حسابية عند قلب المصفوفة  $X'X$  بغية تقدير معاملات الانحذار. والتعبير عن المتغير المستقل كإنحراف عن متوسطه يخفف كثير من الخطية المتعددة والتي سوف نتناولها في الفصل التاسع. سوف نوضح ذلك في المثال التالي:

مثال (٦-٦)

توضح البيانات في جدول (٦-١٥) عدد الوحدات المنتجة في مصنع  $x$  على مدى 18 أسبوعا وكذلك التكاليف المقدرة للوحدة  $y$  على مستوى الإنتاج المناظر.

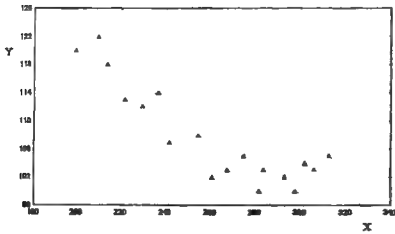
والمطلوب: إيجاد معادلة الانحذار من الدرجة الثانية باستخدام انحرافات قيم  $x$ .

جدول (٦-١٥)

x	y	x	y
242	107	200	120
255	108	210	122
261	102	237	114
268	103	284	103
275	105	293	102
282	100	298	100
222	113	302	104
214	118	306	103
230	112	313	105

الحل

شكل الانتشار موضح في شكل (٦-١١).



شكل (٦-١١)

معادلة الانحدار المقدره سوف تكون :

$$\hat{y} = b_0 + b_1x + b_2x^2. \quad (٤-٦)$$

وبدلالة انحرافات  $x$  عن المتوسط الحسابي سوف تكون:

$$\hat{y}' = b'_0 + b'_1z + b'_2z^2. \quad (٥-٦)$$

$$\bar{y}=107.8333, \quad \bar{x}=260.6667.$$

حيث :

وبما أن:

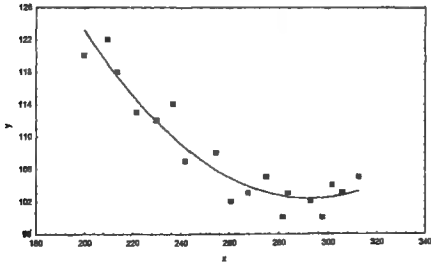
قيم  $b'_0, b'_1, b'_2$  سوف تكون:

$$b'_2 = .002411, \quad b'_1 = -.154964, \quad b'_0 = -2.947.$$

وبالتعويض بقيم  $b'_0, b'_1, b'_2$  في المعادلة (٥-٦) نحصل على معادلة الانحدار المقدره:

$$\hat{y} = 309.1 + .002411x - 1.411925x^2.$$

والموضحة بيانياً في شكل (١٢-٦) مع شكل الانتشار.



شكل (١٢-٦)

(٦-١-٦) كثيرات الحدود المتعامدة

**Orthogonal Polynomials**

تستخدم كثيرات الحدود المتعامدة لتحويل المتغيرات المستقلة (الحدود) في نموذج الانحدار لكثيرات الحدود وذلك بهدف.

- تحديد درجة المعادلة وتقدير معاملاتها.
- تقادي مشكلة الارتباط الخطي المتعدد.

ويشترط عند استخدامها أن تكون قيم المتغير المستقل على أبعاد متساوية  
فعلى سبيل المثال 4, 8, 12 حيث المسافة بين كل قيمة والأخرى  $D=4$ .

**تحديد درجة المعادلة:**

سوف نستخدم الطريقة الامامية لتحديد درجة النموذج ، أي إضافة حدود الدرجات العليا واحدا بعد الآخر إلى أن تصبح الإضافة غير معنوية لمرتين متتاليتين. إن الخطوة الأولى هي الإندار الخطي البسيط ثم النموذج من الدرجة الثانية ثم التكعيبية وهكذا. هذا ولانقف عن الاختبار حتى نحصل على قيمة لـ  $F$  غير معنوية لمرتين متتاليتين ، ولو فرضنا أن النموذج من الدرجة الرابعة مثلا هو الذي يمثل البيانات وكانت الدرجة الثانية  $x^2$  غير معنوية فلن هذا لايعني أن تحذف  $x^2$  من المعادلة.

بفرض النموذج المقترح هو:

$$Y_j = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k + \epsilon_j \quad (٦-٦)$$

الاعده في المصفوفة  $X$  لا تكون متعامدة وأكثر من ذلك عندما نرغب في زيادة درجة كثيرات الحدود وذلك بإضافة الحد  $\beta_{k+1} x^{k+1}$  لابد من إعادة حساب  $(X'X)$  وإعادة تقدير المعاملات ذات الرتبة الأقل وهي  $b_0, b_1, \dots, b_k$  والتي سوف تتغير. إن استخدام كثيرات الحدود المتعامدة تعطينا النموذج التالي:

$$Y_j = \alpha_0 g_0(x_j) + \alpha_1 g_1(x_j) + \alpha_2 g_2(x_j) + \dots + \alpha_k g_k(x_j) + \epsilon_j \quad (٧-٦)$$

حيث  $g_i(x_j)$  هي كثيرات الحدود من الرتبة  $i$  وتستخرج من الجدول في ملحق

(٧) . تحقق  $g_i(x_j)$  الشروط التالية:

$$\sum_{i=1}^n g_r(x_i) g_s(x_i) = 0, \quad r \neq s,$$

$$r, s = 0, 2, \dots, k,$$

$$g_0(x_i) = 1$$

ويلاحظ أن القيم  $g_i(x_j)$  دالة في قيم المتغير المستقل. إن  $Y_j$  في (٧-٦) تساوي  $Y_j$  في (٦-٦) إلا أن  $\alpha_i$  لا تساوي  $\beta_i$  وإن  $g_i(x_j)$  لا تعادل  $x_j$ . وعليه فإن تفسير النتائج لا يمكن إستنتاجها مباشرة من النموذج (٧-٦) ولا بد من تحويل النموذج المقدر في (٧-٦) إلى النموذج المقدر في (٦-٦). أيضا فإن استخدام النموذج (٧-٦) لا يعطينا جميع الإحصاءات التي نحصل عليها باستخدام النموذج (٦-٦). ولكن استخدام النموذج (٧-٦) يؤدي إلى سهولة حساب مجموع المربعات لكل درجة لاختبارها مباشرة في تحديد درجة النموذج ويمكن استخدام جداول كثيرات الحدود على المشاهدات  $Y_j$  الغير مكرره ، أو على مجموع المشاهدات  $Y_j$  المكرره مع الأخذ في الاعتبار التصحيح لعدد المكررات  $r$ . كما سنوضح فيما بعد.

بصيغة المصفوفات فإن النموذج (٧-٦) يصبح  $Y = X\alpha + \epsilon$  حيث المصفوفة  $X$  تكون:

$$X = \begin{bmatrix} g_0(x_1) & g_1(x_1) & \dots & g_k(x_1) \\ g_0(x_2) & g_1(x_2) & \dots & g_k(x_2) \\ \vdots & \vdots & & \vdots \\ g_0(x_n) & g_1(x_n) & \dots & g_k(x_n) \end{bmatrix}$$

وللتسهيل سوف نضع :

$$g_0(x_j) = 1, \quad j = 1, 2, \dots, n, \quad g_i(x_j) = z_{ij}, \quad i = 1, 2, \dots, k.$$

حيث  $g_0(x_j)$  كثيرات الحدود من الدرجة صفر أي أن  $Z$  تصبح:

$$Z = \begin{bmatrix} 1 & z_{11} & z_{21} & \dots & z_{k1} \\ 1 & z_{12} & z_{22} & \dots & z_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & z_{1n} & z_{2n} & & z_{kn} \end{bmatrix}$$

وبما أن المصفوفة  $Z$  لها أعمدة متعامدة فإن  $Z'Z$  تصبح:

$$Z'Z = \begin{bmatrix} n & 0 & \dots & 0 \\ 0 & \sum_{j=1}^n z_{1j}^2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sum_{j=1}^n z_{kj}^2 \end{bmatrix}$$

مقدرات المربعات الصغرى للمتجه  $\alpha$  يمكن الحصول عليها من  $(Z'Z)^{-1}Z'y$  حيث:

$$\hat{\alpha}_i = \frac{\sum_{j=1}^n z_{ij} y_j}{\sum_{j=1}^n z_{ij}^2}, \hat{\alpha}_0 = \bar{y} = \frac{\sum_{j=1}^n y_j}{n}.$$

مجموع مربعات البواقي ستكون:

$$SSE = SYY - \sum_{i=1}^k \hat{\alpha}_i \left[ \sum_{j=1}^n z_{ij} y_j \right].$$

إن مجموع مربعات الانحدار لأي معلمة في النموذج لا يعتمد على المعالم الأخرى في النموذج. مجموع مربعات الانحدار سوف تكون:

$$\begin{aligned} SSR(\alpha_i) &= \hat{\alpha}_i \sum_{j=1}^n z_{ij} y_j \\ &= \frac{(\sum z_{ij} y_j)^2}{\sum_{j=1}^n z_{ij}^2} \end{aligned}$$

جدول تحليل التباين باستخدام كثيرات الحدود المتعامدة عند عدم وجود تكرار لقيم x معطى في جدول (١٦-٦).

جدول (١٦-٦)

S.O.V	df	SS
الكلي	n-1	SY Y
$\beta_1   \beta_0$	1	$SSR(\alpha_1) = \frac{(\sum z_{1j} y_j)^2}{\sum_{j=1}^n z_{1j}^2}$
خطأ $(\beta_1   \beta_0)$	n-2	$SSE(\alpha_1) = SY Y - SSR(\alpha_1)$
$\beta_2   \beta_1, \beta_0$	1	$SSR(\alpha_2) = \frac{(\sum z_{2j} y_j)^2}{\sum_{j=1}^n z_{2j}^2}$
خطأ $(\beta_2   \beta_1, \beta_0)$	n-3	$SSE(\alpha_2) = SSE(\alpha_1) - SSR(\alpha_2)$
$\beta_3   \beta_2, \beta_1, \beta_0$	1	$SSR(\alpha_3) = \frac{(\sum z_{3j} y_j)^2}{\sum_{j=1}^n z_{3j}^2}$
خطأ $(\beta_3   \beta_2, \beta_1, \beta_0)$	n-4	$SSE(\alpha_3) = SSE(\alpha_2) - SSR(\alpha_3)$

لحساب  $SSR(\alpha_i)$  ،  $\hat{\alpha}_i$  يستحسن لتسهيل الحساب عمل جدول كالمعطى في جدول (١٧-٦)



جدول (١٧-٦)

درجة المعادلة	قيم $z_j$	$\sum z_{ij}y_j$	$\sum z_{ij}^2$	$SSR(\alpha_i)$
خطية	قيم $z_{1j}$ من الجدول في ملحق (٧)	.	.	$SSR(\alpha_1)$
تربيعيه	قيم $z_{2j}$ من الجدول في ملحق (٧)	.	.	$SSR(\alpha_2)$
تكعيبيه	قيم $z_{3j}$ من الجدول في ملحق (٧)	.	.	$SSR(\alpha_3)$
:	.	.	.	:

ويجب تحويل المعادلة المقدرة بطريقة كثيرات الحدود إلى معادلة الانحدار العادية وذلك بالتعويض عن  $z_1, z_2, \dots$  بقيمتها التي هي دالة في  $x$ . فمثلا إذا كان نموذج الانحدار من الدرجة الخامسة فإننا نضع  $g_0(x) = 1$  ونعوض عن القيم التالية:  $z_1, z_2, z_3, z_4, z_5$

$$z_1 = g_1(x_j) = \left[ \frac{x_j - \bar{x}}{D} \right] \lambda_1,$$

$$z_2 = g_2(x_j) = \left[ \left( \frac{x_j - \bar{x}}{D} \right)^2 - \frac{n^2 - 1}{12} \right] \lambda_2,$$

$$z_3 = g_3(x_j) = \left[ \left( \frac{x_j - \bar{x}}{D} \right)^3 - \left( \frac{x_j - \bar{x}}{D} \right) \left( \frac{3n^2 - 7}{20} \right) \right] \lambda_3,$$

$$z_4 = g_4(x_j) = \left[ \left( \frac{x_j - \bar{x}}{D} \right)^4 - \left( \frac{x_j - \bar{x}}{D} \right)^2 \left( \frac{3n^2 - 13}{14} \right) + \frac{3}{560} (n^4 - 10n^2 + 9) \right] \lambda_4,$$

$$z_5 = g_5(x_j) = \left[ \left( \frac{x_j - \bar{x}}{D} \right)^5 - \left( \frac{x_j - \bar{x}}{D} \right)^3 \frac{5(n^2 - 7)}{18} + \left( \frac{x_j - \bar{x}}{D} \right) \left( \frac{15n^4 - 230n^2 + 407}{1008} \right) \right] \lambda_5.$$

حيث أن  $D$  هو الفرق بين مستويات  $x$  و  $\lambda_5$  تستخرج من الجدول في الملحق (٧) من العمود الأخير عندما يكون عدد مستويات  $x$  تساوي  $k$ . ويجب ان نعلم ان  $\lambda_5$  و  $i = 1, 2, \dots, 5$  تختار بحيث ان كثرات الحدود تكون قيم موجبه. ويمكن الحصول عليها في الحالات التي تكون فيها المسافه بين مستويات  $x$  غير متساوية وذلك بالرجوع إلى (Seber (1977).

مثال (٦-٧)

لازواج المشاهدات المعطاه في جدول (٦-١٨) اوجد معادلة الاتجاه المقدره من الدرجة الثانية باستخدام أسلوب كثرات الحدود.

جدول (٦-١٨)

x	0	1	2	3	4	5	6	7	8	9
y	9.1	7.3	3.2	4.6	4.8	2.9	5.7	7.1	8.8	10.2

الحل

بما أن المسافه بين قيم  $x$  متساوية أى أن الفرق بين قيمتين متتاليتين من  $x$  يساوي 1 لذلك يمكن استخدام طريقة كثرات الحدود المتعامدة لتحديد درجة النموذج الذي يمثل البيانات المعطاه في جدول (٦-١٨). بما أن  $n=10$  فيمكن نظريا حساب معادلات إلى حد الدرجة التاسعة  $9=(10-1)$  ولكن عمليا يستحسن ان لا تزيد درجة المعادلة عن الدرجة الخامسة. ولإيجاد قيم  $z_j$  التابعة للدرجة الاولى والثانية والثالثة والرابعة والخامسة يستخدم الجدول في ملحق (٧) حيث:

$$Z_{1j}, Z_{2j}, Z_{3j}, Z_{4j}, Z_{5j}$$

هم الدرجة الخطية والدرجة التربيعية و الدرجة التكعيبية و الدرجة الرابعة و

الدرجة الخامسة على التوالي. كما هي معطاة في جدول (٦-١٩) حيث  $\sum_{j=1}^n Z_{1j}y_j$

مثلا هي:

$$\begin{aligned}\sum Z_{1j}y_j &= (-9)(9.1) + (-7)(7.3) + (-5)(3.2) + (-3)(4.6) \\ &+ (-1)(4.8) + (1)(2.9) + (3)(5.7) \\ &+ (5)(7.1) + (7)(8.8) + (9)(10.2) = 41.3.\end{aligned}$$

وأن  $\sum Z_{1j}^2 = 330$  وتستخرج من الجدول في ملحق (٧) في العمود قبل الأخير من الجدول على اليمين. ويمكن حسابها بدون جدول كالتالي:

$$(-9)^2 + (-7)^2 + (-5)^2 + \dots + (9)^2 = 330.$$

وتحسب قيم  $\alpha$  كالآتي:

$$\hat{\alpha}_i = \frac{\sum Z_{ij}y_j}{\sum Z_{ij}^2}.$$

فمثلا:

$$\hat{\alpha}_1 = \frac{41.3}{330} = 0.1252,$$

$$\hat{\alpha}_2 = \frac{76}{132} = 0.5758,$$

$$\hat{\alpha}_3 = \frac{-122.6}{8580} = -0.014289.$$

$$\hat{\alpha}_4 = \frac{-12.4}{2860} = -0.004335.$$

$$\hat{\alpha}_5 = \frac{-9.8}{780} = -0.0125.$$

ولإيجاد مجموع مربعات الانحدار التي تعود للدرجة الخطية والتربيعية والتكعيبية والرابعة والخامسة نتبع ما يأتي:

$$SSR(\alpha_1) = \frac{\left( \sum_{j=1}^n z_{1j} y_{1j} \right)^2}{\sum_{j=1}^n z_{1j}^2} = \frac{(41.3)^2}{330} = 5.1688,$$

$$SSR(\alpha_2) = \frac{\left( \sum_{j=1}^n z_{2j} y_{2j} \right)^2}{\sum_{j=1}^n z_{2j}^2} = \frac{(76)^2}{132} = 43.7576,$$

$$SSR(\alpha_3) = \frac{\left( \sum z_{3j} y_{3j} \right)^2}{\sum z_{3j}^2} = \frac{(-122.6)^2}{8580} = 1.7518,$$

$$SSR(\alpha_4) = \frac{\left( \sum z_{4j} y_{4j} \right)^2}{\sum z_{4j}^2} = \frac{(-12.4)^2}{2860} = 0.0538,$$

$$SSR(\alpha_5) = \frac{\left( \sum z_{5j} y_{5j} \right)^2}{\sum z_{5j}^2} = \frac{(-9.8)^2}{780} = 0.1231.$$

جدول (١٩-٦)

درجة المعادلة	y <sub>1</sub>	y <sub>2</sub>	y <sub>3</sub>	y <sub>4</sub>	y <sub>5</sub>	y <sub>6</sub>	y <sub>7</sub>	y <sub>8</sub>	y <sub>9</sub>	y <sub>10</sub>	$\sum y_j$	$\sum z_{ij}^2$	SSR <sub>j</sub>
خطية z <sub>1j</sub>	9.1	7.3	3.2	4.6	4.8	2.9	5.7	7.1	8.8	10.2	41.3	330	5.1688
تربيعية z <sub>2j</sub>	6	2	-1	-3	-4	-4	-3	-1	2	6	76	132	43.7576
تكعيبية z <sub>3j</sub>	-42	14	35	31	12	-12	-31	-35	-14	42	-122.6	8580	1.7518
رابعة z <sub>4j</sub>	18	-22	-17	3	18	18	3	-17	-22	18	-12.4	2860	0.0538
خامسة z <sub>5j</sub>	-6	14	-1	-11	-6	6	11	1	-14	6	-9.8	780	0.1231

ولإيجاد مجموع مربعات الخطأ التابع لكل درجة نتبع مايلي:

• مجموع مربعات الخطأ من الدرجة الأولى هو:

$$SSE(\alpha_1) = SYY - SSR(\alpha_1) \\ = 57.561 - 5.1688 = 52.4222,$$

$$SSY = \sum y_j^2 - \frac{(\sum y_j)^2}{n} \\ = 463.33 - \frac{(63.7)^2}{10} = 57.561.$$

• مجموع مربعات الخطأ من الدرجة الثانية هو:

$$SSE(\alpha_2) = SSE(\alpha_1) - SSR(\alpha_2) \\ = 52.4222 - 43.7576 = 8.6646,$$

• مجموع مربعات الخطأ من الدرجة الثالثة هو:

$$SSE(\alpha_3) = SSE(\alpha_2) - SSR(\alpha_3) \\ = 8.6646 - 1.7518 = 6.9128,$$

• مجموع مربعات الخطأ من الدرجة الرابعة هو:

$$SSE(\alpha_4) = SSE(\alpha_3) - SSR(\alpha_4) \\ = 6.9128 - 0.0538 = 6.859,$$

• مجموع مربعات الخطأ من الدرجة الخامسة هو:

$$\begin{aligned} SSE(\alpha_5) &= SSE(\alpha_4) - SSR(\alpha_5) \\ &= 6.859 - 0.1231 = 6.7359. \end{aligned}$$

جدول تحليل التباين بإستخدام كثيرات الحدود عند عدم وجود تكرار لقيم  $y$  معطى في جدول (٢٠-٦).

جدول (٢٠-٦)

S.O.V	df	SS	MS	F
المجموع	9	57.561		
الدرجة الاولى	1	5.1688	5.1688	
خطأ الدرجة الاولى	8	52.4222	6.549	0.78879
الدرجة الثانية	1	43.7576	43.7576	35**
خطأ الدرجة الثانية	7	8.6646	1.2378	
الدرجة الثالثة	1	1.7518	1.7518	1.52
خطأ الدرجة الثالثة	6	6.9128	1.152	
الدرجة الرابعة	1	0.0538	0.0538	<1
خطأ الدرجة الرابعة	5	6.859	0.3718	
الدرجة الخامسة	1	0.1231	0.1232	<1
خطأ الدرجة الخامسة	4	6.7359	0.68	

ثم نختبر كل درجة بإستخدام الخطأ التابع لها. فأول اختبار هو الذي يتعلق بالدرجة الخطية أي:

$$H_0 : \alpha_1 = 0.$$

القيمة المحسوبة للإحصاء F هي:

$$F = \frac{MSR(\alpha_1)}{MSE(\alpha_1)} = \frac{5.1688}{6.549} = 0.789.$$

وبما أن قيمة F المحسوبة (0.789) أقل من الواحد الصحيح لذا فإننا نقبل  $H_0 : \alpha_1 = 0$ . لاختبار الدرجة التربيعية فإن القيمة المحسوبة للإحصاء F هي:

$$F = \frac{MSR(\alpha_2)}{MSE(\alpha_2)} = \frac{43.7576}{1.2378} = 35$$

وبما أن قيمة F المحسوبة (35) تزيد عن القيمة الجدولية  $F_{0.05}(1,7)=5.59$  فإن  $\alpha_2 \neq 0$ . أي أن الدرجة التربيعية تختلف عن الصفر. ثم نختبر الدرجة التكعيبية والدرجة الرابعة والدرجة الخامسة وبما أنهما لا يختلفان عن الصفر لذلك فإن المعادلة التربيعية هي التي تمثل البيانات خير تمثيل. ومما يجدر الإشارة إليه أننا وجدنا الدرجة الأولى (الخطية) غير معنوية ولكن هذا لا يعني أن نحذفها من النموذج لأن الدرجة التربيعية معنوية.

التقديرات للنموذج التربيعي يمكن الحصول عليها كالتالي:

$$\hat{\alpha}_0 = \bar{y} = \frac{63.7}{10} = 6.37,$$

$$\hat{\alpha}_1 = \frac{\sum z_{1j} y_j}{\sum z_{1j}^2} = \frac{41.3}{330} = 0.125,$$

$$\hat{\alpha}_2 = \frac{\sum z_{2j} y_j}{\sum z_{2j}^2} = \frac{76}{132} = 0.5758,$$

أي أن معادلة الانحدار المقطرة لكثيرات الحدود تصبح :

$$\hat{y} = 6.37 + 0.125z_1 + 0.5758z_2. \quad (٨-٦)$$

لتحويل هذه المعادلة إلى معادلة الاتحاد الأصلية نعوض عن  $z_1$  بـ

$$z_1 = \lambda_1 \left[ \frac{x - \bar{x}}{D} \right]$$

حيث  $D=1$ . في هذا المثال  $\bar{x} = 4.5$  ومن الجدول في ملحق (٧) فإن  $\lambda_1 = 2$  وذلك عندما  $k=10$  من الجدول.

وعن  $z_2$  بـ:

$$z_2 = \lambda_2 \left\{ \left[ \frac{x - \bar{x}}{D} \right]^2 - \left( \frac{n^2 - 1}{12} \right) \right\}.$$

من الجدول في ملحق (٧) فإن  $\lambda_2 = 1/2$  وذلك عند  $k=10$  من الجدول. إذن :

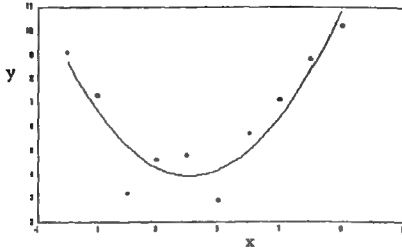
$$\begin{aligned} z_1 &= 2(x - 4.5), \\ z_2 &= 1/2 \left[ (x - 4.5)^2 - (99/12) \right], \end{aligned}$$

وبالتعويض عن  $z_1$ ,  $z_2$  بما يعادلها اعلاه نجد أن المعادلة (٦-٨) بعد التبسيط تصبح :

$$\hat{y} = 8.698 - 2.341x + 0.288x^2.$$

والممثلة بيانيا مع شكل الانتشار في شكل (٦-١٣).





شكل (٦-١٣)

مثال (٦-٨)

في تجربة أجريت لتحديد أثر درجة حرارة التخزين على فاعلية أحد المضادات الحيوية أخذت 15 عينة من المضاد الحيوي وتم تقسيمها عشوائياً إلى 5 مجموعات عرضت كل منها لدرجات الحرارة التالية:  $10^{\circ}, 30^{\circ}, 50^{\circ}, 70^{\circ}, 90^{\circ}$  درجة مئوية، وبعد شهر من التخزين تم اختبار الفاعلية والحصول على النتائج المعطاه في جدول (٦-٢١). والمطلوب إيجاد العلاقة بين الاستجابة  $Y$  ومستويات  $x$  و تحديد درجة نموذج الانحدار التي توافق البيانات في جدول (٦-٢١) باستخدام اختبار نقص التوافق وإيجاد معادلة الانحدار المقدره.

جدول (٦-٢١)

	$10^{\circ}$	$30^{\circ}$	$50^{\circ}$	$70^{\circ}$	$90^{\circ}$
	62	26	16	10	13
	55	36	15	11	11
	57	31	23	18	9
$y_i$	174	93	54	39	33
$\bar{y}_j$	58	31	18	13	11

الحل

تحديد درجة نموذج الانحدار :-

لاحظ أولاً أن المسافة بين قيم  $x_i$  (درجات الحرارة)  $20^\circ$  وأن عدد قيم  $x$  المميزة هي 5 ( $k=5$ ) وكل منهما قد تكرر ثلاثة مرات ( $r=3$ ). سوف نرمز للملاحظة رقم  $j$  التابعة للمجموعة  $i$  بـ  $y_{ij}$  ونرمز لمجموع المشاهدات للمجموعة  $i$  بالرمز  $y_{i.}$  ونرمز للمجموع الكلي بالرمز  $y_{..}$ . وتتلخص طريقة كثيرات الحدود المتعامدة لتحديد درجة نموذج الانحدار عند وجود تكرار لقيم  $x$  فيما يأتي:

١- إيجاد مجموع المربعات الكلية  $SY Y$  وهو:

$$\begin{aligned} SY Y &= \sum \sum y_{ij}^2 - \frac{(y_{..})^2}{rk} \\ &= (62)^2 + (55)^2 + (57)^2 + (26)^2 + \dots + (9)^2 - \frac{(393)^2}{(3)(5)} \\ &= 4680.4. \end{aligned}$$

٢- إيجاد مجموع المربعات للمجاميع  $SSB.G$  وهو:

$$\begin{aligned} SSB.G &= \frac{\sum y_{i.}^2}{r} - \frac{(y_{..})^2}{rk} \\ &= \frac{(174)^2 + (93)^2 + \dots + (33)^2}{3} \\ &\quad - \frac{(393)^2}{(3)(5)} = 4520.4. \end{aligned}$$

٣- إيجاد مجموع المربعات للخطأ:

$$\begin{aligned} SSE &= SY Y - SSB.G. \\ &= 4680.4 - 4520.4 = 160.0. \end{aligned}$$

لاحظ أن هذا هو مجموع مربعات الخطأ الخالص  $SSPE$ .

٤- الآن نحسب مجموع المربعات الذي يعود للدرجة الخطية:

$$SSR(\alpha_1) = \frac{(\sum z_{i1} \bar{y}_{i.})^2}{\sum z_{i1}^2 / r}$$

حيث قيم  $\sum_{i=1}^k z_{ij} \bar{y}_{i.}$  تحسب كالتالي:

$$\sum z_{i1} \bar{y}_{i.} = (-2)(58) + (-1)(31) + (0)(18) + (1)(13) + (2)(11) = -112$$

$$\sum z_{i2} \bar{y}_{i.} = (2)(58) + (-1)(31) + (-2)(18) + (-1)(13) + (2)(11) = 58$$

$$\sum z_{i3} \bar{y}_{i.} = (-1)(58) + (2)(31) + (0)(18) + (-2)(13) + (1)(11) = -11$$

$$\sum z_{i4} \bar{y}_{i.} = (1)(58) + (-4)(31) + (6)(18) + (-4)(13) + (1)(11) = 1.$$

جدول تحليل التباين موضح في جدول (٦-٢٢).

جدول (٦-٢٢)

الدرجة	10°	30°	50°	70°	90°	$\sum z_{ij} \bar{y}_{i.}$	$\sum z_{ij}^2 / r$	SSR( $\alpha_j$ )
	58	31	18	13	11			
الأولي	-2	-1	0	1	2	-112	(10)/3	3763.20
الثانية	2	-1	-2	-1	2	58	(14)/3	720.86
الثالثة	-1	2	0	-2	1	-11	(10)/3	36.30
الرابعة	1	-4	6	-4	1	1	(70)/3	0.04
							SSB.G	4520.40

مجموع المربعات المقابل للاتجاه لأي درجة  $j$  حيث  $j = 1, 2, 3, 4$  يتم حسابه عن طريق الصيغة التالية:

$$SSR(\alpha_j) = \frac{(\sum z_{ij} \bar{y}_{i.})^2}{\sum z_{ij}^2 / r}, j = 1, 2, 3, 4.$$

وذلك عند استخدام المتوسط في حساب  $SSR(\alpha_j)$ ، أما عند استخدام المجموع في حساب  $SSR(\alpha_j)$  فإن  $SSR(\alpha_j)$  تصبح كالتالي:

$$SSR(\alpha_j) = \frac{(\sum z_{ij} \bar{y}_i)^2}{r \sum z_{ij}^2}$$

مجموع المربعات المقابل للدرجة الأولى باستخدام المتوسطات هو:

$$\begin{aligned} SSR(\alpha_1) &= \frac{(\sum z_{i1} \bar{y}_i)^2}{\sum z_{i1}^2 / r} \\ &= (-112)^2 / [(-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2 / 3] \\ &= (-112)^2 / (10/3) \\ &= 3763.2, \end{aligned}$$

مجموع المربعات المقابل للدرجة الثانية هو:

$$\begin{aligned} SSR(\alpha_2) &= \frac{(\sum z_{i2} \bar{y}_i)^2}{\sum z_{i2}^2 / r} \\ &= (58)^2 / [(2)^2 + (-1)^2 + (-2)^2 + (-1)^2 + (2)^2 / 3] \\ &= (58)^2 / (14/3) \\ &= 720.86, \end{aligned}$$

مجموع المربعات المقابل للدرجة الثالثة هو:

$$\begin{aligned} SSR(\alpha_3) &= \frac{(\sum z_{i3} \bar{y}_i)^2}{\sum z_{i3}^2 / r} \\ &= (-11)^2 / [(-1)^2 + (2)^2 + (0)^2 + (-2)^2 + (1)^2 / 3] \\ &= (-11)^2 / (10/3) \\ &= 36.3, \end{aligned}$$

مجموع المربعات المقابل للدرجة الرابعة هو:

$$\begin{aligned} SSR(\alpha_4) &= \frac{(\sum z_{i4} \bar{y}_i)^2}{\sum z_{i4}^2 / r} \\ &= (1)^2 / [(1)^2 + (-4)^2 + (6)^2 + (-4)^2 + (1)^2 / 3] \\ &= (1)^2 / (70/30) \\ &= 0.04. \end{aligned}$$

لاختبار الفرضية  $H_0: \beta_1 = 0$  أو  $H_0: \alpha_1 = 0$  فإننا نستخدم الإحصاء F حيث أن :

$$F = \frac{SSR(\alpha_1)}{MSE}.$$

القيمة المحسوبة للإحصاء F هي:

$$F = \frac{3763.2}{160.0/10} = 235.2.$$

وبما أن قيمة F المحسوبة تزيد عن قيمة F الجدولية عند مستوى معنوية  $\alpha = 0.01$  والتي تساوي  $F_{0.01}[1,10] = 10.04$  فإننا نرفض  $H_0$  أي نقبل  $H_0: \alpha_2 = 0$  أو  $H_1: \beta_1 \neq 0$ . ثم نختبر فرض العدم  $H_0: \alpha_2 = 0$  أو  $H_1: \beta_2 = 0$  باستخدام الإحصاء F حيث:

$$F = \frac{SSR(\alpha_2)}{MSE}.$$

القيمة المحسوبة للإحصاء F هي:

$$F = \frac{720.86}{160.0/10} = 45.05.$$

بما وأن قيمة F المحسوبة تزيد عن F الجدولية عند مستوى معنوية  $\alpha = 0.01$  والتي تساوي  $F_{0.01}[1,10] = 10.04$  فإننا نرفض  $H_0$  أي نقبل  $H_1: \beta_2 \neq 0$ . ثم نختبر فرض العدم  $H_0: \alpha_3 = 0$  أو  $H_1: \beta_3 = 0$  باستخدام الإحصاء F حيث:

$$F = \frac{SSR(\alpha_3)}{MSE} .$$

القيمة المحسوبة للإحصاء F هي:

$$F = \frac{36.3}{160.0/10} = 2.27.$$

وبما أن قيمة F المحسوبة أصغر من قيمة F الجدولية عند مستوى معنوية  $\alpha = 0.01$  والتي تساوي  $F_{0.01}[1,10] = 10.04$  فإننا نقبل  $H_1 : \beta_3 = 0$ . وعلى ذلك نتوقف عند هذا الحد، أما إذا رفضنا فرض العدم عند هذه الخطوة فإننا نختبر الدرجة الرابعة. ولما كان قصور التوفيق غير معنوي عند إضافة الدرجة الثالثة فإنه يمكننا استنتاج أن فاعلية المضاد الحيوي تأخذ شكل معادلة من الدرجة الثانية في درجة حرارة التخزين.

#### إيجاد منحني الاستجابة المقدر:

منحني الاستجابة المقدر يأخذ الشكل التالي:

$$\hat{Y}_i = \hat{\alpha}_0 + \hat{\alpha}_1 z_{i1} + \hat{\alpha}_2 z_{i2}.$$

حيث:

$$\hat{\alpha}_0 = \bar{y}_{..} = \frac{393}{15} = 26.2,$$

$$\hat{\alpha}_1 = \frac{\sum z_{i1} \bar{y}_i}{\sum z_{i1}^2} = \frac{-112}{10} = -11.2,$$

$$\hat{\alpha}_2 = \frac{\sum z_{i2} \bar{y}_i}{\sum z_{i2}^2} = \frac{58}{14} = 4.14.$$

وعلى ذلك:

$$y = \hat{\alpha}_0 + \hat{\alpha}_1 z_{i1} + \hat{\alpha}_2 z_{i2}$$

$$= 26.2 + (-11.2)z_{i1} + (4.14)z_{i2}.$$

بنفس الطريقة يمكن حساب القيم المتوقعة عند درجات الحرارة المختلفة كما هو موضح في جدول (٦-٢٣).

جدول (٦-٢٣)

درجة الحرارة	10°	30°	50°	70°	90°
المتوسط الشاهد	58	31	18	13	11
المتوسط المتوقع	56.88	33.3	17.92	10.86	12.08

إذا كان المطلوب هو إيجاد معادلة الانحدار المقدره بإستخدام قيم  $x$  المختلفة فيتم حساب معادلة الدرجة الثانية كما يلي:

$$\hat{y}_i = b_0 + b_1 x_i + b_2 x_i^2$$

$$= \hat{\alpha}_0 + \hat{\alpha}_1 z_{i1} + \hat{\alpha}_2 z_{i2}$$

$$= \bar{y}_{..} + \hat{\alpha}_1 \left[ \frac{x_i - \bar{x}}{D} \right] \lambda_1 + \hat{\alpha}_2 \left( \left[ \frac{x_i - \bar{x}}{D} \right]^2 - \left[ \frac{n^2 - 1}{12} \right] \right) \lambda_2.$$

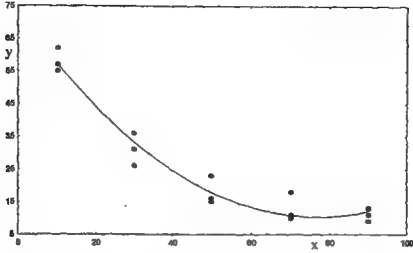
$$= 26.2 + (-11.2) \left[ (x_i - 50) / 20 \right] (1)$$

$$+ (4.14) \left[ \left[ (x_i - 50) / 20 \right]^2 - (5^2 - 1) / 12 \right] (1), \lambda_1 = 1, \lambda_2 = 1.$$

و بعد ترتيب وتبسيط المعادلة السابقة نحصل على العلاقة من الدرجة الثانية على الشكل:

$$\hat{y}_i = 71.81 - 1.596x_i + 0.01036x_i^2 \quad (٦-٩)$$

والمنبئة بيانيا في شكل (٦-١٤)



شكل (٦-١٤)

حيث  $\hat{y}$  الفاعلية المتوقعة و  $x_i$  درجة حرارة التخزين ، ويمكن للباحث وضع تقرير إحصائي من النتائج سالفة الذكر كما يلي:

خلال مدة شهر من التخزين في درجات الحرارة المختلفة وجد أن هناك تساثير معنوي عالي لدرجات الحرارة المختلفة على فاعلية المضاد والتي تقل مع الزيادة في درجة الحرارة. وقد أثبت التحليل الإحصائي أنه يمكن وصف الفاعلية على مدى درجات حرارة تتراوح بين  $10^\circ$  و  $90^\circ$  بالمعادلة التالية:

$$\hat{y} = 71.81 - 1.596x + 0.01x^2 .$$

حيث  $\hat{y}$  الفاعلية المتوقعة و  $x$  درجات حرارة التخزين خلال 30 يوماً.

وفي الحقيقة أن الميزة الأساسية لاستخدام كثيرات الحدود المتعامدة هي سهولة حذف وإضافة أي درجة في النموذج بدون أن يؤثر ذلك في تقديرات المعاملات الأخرى لنموذج الانحدار . هذا ويمكن استخدام الحزم الجاهزة وذلك باستخدام الحاسب الآلي باستخدام طريقة المربعات الصغرى وذلك للحصول على المعادلة السابقة لمهولة الحساب.

وفي النهاية فإن هذا الموضوع يرتبط أكثر بتصميم التجارب ويعمل واحد عند إيجاد منحنى الإستجابة ولمزيد من التفاصيل يمكن الرجوع إلى كتاب المؤلفة تصميم وتحليل التجارب والموجود في المراجع.



(٢-٦) نماذج انحدار كثيرات الحدود - متغيرين مستقلين

الأسلوب المتبع لتوفيق نموذج انحدار كثيرات الحدود يمكن تعميمه للحالة التي يكون فيها أكثر من متغير مستقل واحد. بفرض أن لدينا متغير الإستجابة  $Y$  مع متغيرين مستقلين ( $k=2$ ) ونموذج انحدار من الدرجة الثانية على الصورة:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \epsilon \quad (١٠-٦)$$

يسمى النموذج (١٠-٦) بنموذج انحدار غير مستقيم متعدد لمتغيرين مستقلين حيث الحد  $x_1 x_2$  يمثل التداخل (التفاعل) بين المتغيرين  $x_1$  و  $x_2$ . أن النموذج (١٠-٦) له عدة أشكال من السطوح أهمها:

• شكل سرج الحصان.

• شكل ملقح سطحين منحدرين .

معادلة الانحدار المقطرة سوف تكون على الشكل التالي:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + b_{12} x_1 x_2 \quad (١١-٦)$$

تتكون البيانات اللازمة لإيجاد المعادلة (١١-٦) من  $n$  من المشاهدات على المتغير التابع  $Y$  ومتغيرين مستقلين  $x_1, x_2$  وتكون البيانات على الشكل التالي:

$$\{(y_j; x_{1j}, x_{2j}), j=1, 2, \dots, n\}.$$

في تلك الحالة فإن  $n$  لابد أن تكون على الأقل تساوي 6. هنا يوجد 6 معالم لابد من تقديرها وبالإضافة إلى ذلك بما أن النموذج يحتوي على حدود تربيعية لكل من المتغيرين فلا بد من توافر على الأقل ثلاثة مستويات من كل متغير مستقل.

يمكن بسهولة إثبات أن معادلات المربعات الصغرى هي:

$$X'X b = X'y.$$

حيث:

$$b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_{11} \\ b_{22} \\ b_{12} \end{bmatrix} \text{ و}$$

$$\begin{bmatrix} n & \sum_{j=1}^n x_{1j} & \sum_{j=1}^n x_{2j} & \sum_{j=1}^n x_{1j}^2 & \sum_{j=1}^n x_{2j}^2 & \sum_{j=1}^n x_{1j}x_{2j} \\ \sum_{j=1}^n x_{1j} & \sum_{j=1}^n x_{1j}^2 & \sum_{j=1}^n x_{1j}x_{2j} & \sum_{j=1}^n x_{1j}^3 & \sum_{j=1}^n x_{1j}x_{2j}^2 & \sum_{j=1}^n x_{1j}^2x_{2j} \\ \sum_{j=1}^n x_{2j} & \sum_{j=1}^n x_{1j}x_{2j} & \sum_{j=1}^n x_{2j}^2 & \sum_{j=1}^n x_{1j}^2x_{2j} & \sum_{j=1}^n x_{2j}^3 & \sum_{j=1}^n x_{1j}x_{2j}^2 \\ \sum_{j=1}^n x_{1j}^2 & \sum_{j=1}^n x_{1j}^3 & \sum_{j=1}^n x_{1j}^2x_{2j} & \sum_{j=1}^n x_{1j}^4 & \sum_{j=1}^n x_{1j}^2x_{2j}^2 & \sum_{j=1}^n x_{1j}^3x_{2j} \\ \sum_{j=1}^n x_{2j}^2 & \sum_{j=1}^n x_{1j}x_{2j}^2 & \sum_{j=1}^n x_{2j}^3 & \sum_{j=1}^n x_{1j}^2x_{2j}^2 & \sum_{j=1}^n x_{2j}^4 & \sum_{j=1}^n x_{1j}x_{2j}^3 \\ \sum_{j=1}^n x_{1j}x_{2j} & \sum_{j=1}^n x_{1j}^2x_{2j} & \sum_{j=1}^n x_{1j}x_{2j}^2 & \sum_{j=1}^n x_{1j}^3x_{2j} & \sum_{j=1}^n x_{1j}x_{2j}^3 & \sum_{j=1}^n x_{1j}^2x_{2j}^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_{11} \\ b_{22} \\ b_{12} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n y_j \\ \sum_{j=1}^n x_{1j}y_j \\ \sum_{j=1}^n x_{2j}y_j \\ \sum_{j=1}^n x_{1j}^2y_j \\ \sum_{j=1}^n x_{2j}^2y_j \\ \sum_{j=1}^n x_{1j}x_{2j}y_j \end{bmatrix}$$

ويمكن تحويل النموذج (٦-١١) إلى نموذج الحدار خطى متعدد به خمسة متغيرات مستقلة وذلك حتى يمكننا إيجاد تقديرات للمعالم في النموذج بطريقة سهلة ويتم ذلك بوضع:

$$x_1 = x_1, x_2 = x_2, x_3 = x_1^2, x_4 = x_2^2, x_5 = x_1x_2.$$

المعادلات الطبقية سوف تكون:

$$X'Xb = X'y.$$

$$X'X = \begin{bmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 & \sum x_4 & \sum x_5 \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 & \sum x_1x_3 & \sum x_1x_4 & \sum x_1x_5 \\ \sum x_2 & \sum x_2x_1 & \sum x_2^2 & \sum x_2x_3 & \sum x_2x_4 & \sum x_2x_5 \\ \sum x_3 & \sum x_3x_1 & \sum x_3x_2 & \sum x_3^2 & \sum x_3x_4 & \sum x_3x_5 \\ \sum x_4 & \sum x_4x_1 & \sum x_4x_2 & \sum x_4x_3 & \sum x_4^2 & \sum x_4x_5 \\ \sum x_5 & \sum x_5x_1 & \sum x_5x_2 & \sum x_5x_3 & \sum x_5x_4 & \sum x_5^2 \end{bmatrix},$$

$$b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix}, X'y = \begin{bmatrix} \sum y_j \\ \sum x_{1j}y_j \\ \sum x_{2j}y_j \\ \sum x_{3j}y_j \\ \sum x_{4j}y_j \\ \sum x_{5j}y_j \end{bmatrix}$$

وعلى ذلك:

$$b = (X'X)^{-1}X'y.$$

مثال (٦-٩)

في عينة عشوائية من الحجم  $n = 31$  تم الحصول على البيانات المعطاه في جدول (٦-٢٤) وذلك لمتغير الاستجابة  $Y$  ومتغيرات مستقلة عددها  $k = 3$ . والمطلوب تقدير معالم نموذج الانحدار تحت فرض النموذج (٦-١١) باستخدام القيم المعيارية.

جدول (٦-٢٤)

رقم المشاهدة	$y$	$x_1$	$x_2$	$x'_1$	$x'_2$
1	60.0	2400	54.5	-1.52428	-.57145
2	61.0	2450	56.0	-1.39535	-.35543
3	65.0	2450	58.5	-1.39535	.00461
4	30.5	2500	43.0	-1.26642	-2.22763
5	63.5	2500	58.0	-1.26642	-.06740
6	65.0	2500	59.0	-1.26642	.07662
7	44.0	2700	52.5	-.75070	-.85948
8	52.0	2700	65.5	-.75070	1.01272
9	54.5	2700	68.0	-.75070	1.37276
10	30.0	2750	45.0	-.62177	-1.93960
11	26.0	2775	45.5	-.55731	-1.86759
12	23.0	2800	48.0	-.49284	-1.50755
13	54.0	2800	63.0	-.49284	.65268
14	36.0	2900	58.5	-.23499	.00461

15	53.5	2900	64.5	-.23499	.86870
16	57.0	3000	66.0	.02287	1.08472
17	33.5	3075	57.0	.21627	-.21141
18	34.0	3100	57.5	.28073	-.13941
19	44.0	3150	64.0	.40966	.79669
20	33.0	3200	57.0	.53859	-.21141
21	39.0	3200	64.0	.53859	.79669
22	53.0	3200	69.0	.53859	1.51677
23	38.5	3225	68.0	.60305	1.37276
24	39.5	3250	62.0	.66752	.50866
25	36.0	3250	64.5	.66752	.86870
26	8.5	3250	48.0	.66752	-1.50755
27	30.0	3500	60.0	1.31216	.22063
28	29.0	3500	59.0	1.31216	.07662
29	26.5	3500	58.0	1.31216	-.06740
30	24.5	3600	58.0	1.57002	-.06740
31	26.5	3900	61.0	2.34360	.36465

بما أن:

$$s_2 = 6.944, \bar{x}_2 = 58.468, s_1 = 387.81, \bar{x}_1 = 2991.13,$$

وعلى ذلك:

$$x'_1 = (x_1 - 2991.13/387.81), x'_2 = (x_2 - 58.468)/6.944,$$

$$x'_3 = (x'_1)^2, x'_4 = (x'_2)^2, x'_5 = x'_1 x'_2,$$

والآن نموذج الانحدار المقدر هو:

$$\hat{y} = 40.27 - 13.40x'_1 + 10.26x'_2 + 2.33x'_3 - 2.34x'_4 + 2.60x'_5.$$

وعلى ذلك إذا كانت:

$$x_1 = 3200, x_2 = 57, x'_1 = 0.539, x'_2 = -0.211, x'_3 = (0.539)^2 = 0.2901,$$

$$x'_4 = (-0.211)^2 = 0.0447, x'_5 = (0.539)(-0.211) = -0.1139.$$

فإن:

$$\begin{aligned}\hat{y} &= 40.27 - (13.40)(0.539) + (10.26)(-0.211) \\ &+ (2.33)(0.2901) - (2.34)(0.0447) \\ &+ (2.60)(-0.1139) = 31.16.\end{aligned}$$

مثال (٦-١٠)

البيانات المعطاه في جدول (٦-٢٥) تمثل نمبة التلوث الذي يحدث على درجات حرارة مختلفة وأزمنة تعقيم خلال تفاعل يرتبط بصناعة مشروب.

جدول (٦-٢٥)

مدة التعقيم $x_2$	درجة الحرارة $x_1$		
	75	100	125
15	14.05	10.55	7.55
	14.93	9.48	6.59
20	16.56	13.63	9.23
	15.85	11.75	8.78
25	22.41	18.55	15.93
	21.66	17.98	16.44

المطلوب: إيجاد معادلة الانحدار المقدره تحت فرض النموذج (٦-١١).

الحل

ويستخدم برنامج SPSS فإن معادلة الانحدار المقدرة سوف تكون:

$$\hat{y} = 56.4668 - 0.36235x_1 - 2.75299x_2 + 0.00081x_1^2 + 0.08171x_2^2 + 0.00314x_1x_2. \quad (١٢-٦)$$

#### اختبارات الفروض:

١- اختبار معنوية الانحدار ككل:

ان هذا الاختبار يستخدم لمعرفة هل ان جميع معاملات الانحدار الجزئية في المعادلة التي تحتوي على متغيرين تساوي صفراً. أي أن

$$H_0: \beta_1 = \beta_2 = \beta_{11} = \beta_{22} = \beta_{12} = 0.$$

سوف نستخدم الإحصاء F على الصورة التالية:

$$F = \frac{MSR(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12} | \beta_0)}{MSE(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12} | \beta_0)}$$

بدرجات حرية  $n-k-1$ .

٢- اختبار يخص معامل انحدار جزئي معين:

فعلى سبيل المثال لاختبار فرض العدم  $H_0: \beta_{12} = 0$  فلنأخذ نستخدم الإحصاء F على الصورة التالية:

$$F = \frac{MSR(\beta_{12} | \beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_0)}{MSE(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12}, \beta_0)}$$

بدرجات حرية  $n-k-1$ .

ويمكن استخدام اختبار t حيث:

$$t = \frac{b_{12}}{s.e.(B_{12})}$$

حيث:

$$s.e.(B_{12}) = \sqrt{MSE c_{66}}$$

و  $c_{66}$  هو آخر عنصر قطري في المصفوفة  $(X'X)^{-1}$ .  
للمثال (١٠-٦) المطلوب إيجاد معادلة الانحدار المقدرة لنموذج الانحدار (١١-٦) واختبار معيولي:

- (أ) هل هناك انحدار معنوي عام؟  
 (ب) هل إضافة  $x_1, x_2$  إلى النموذج سيساعد على التنبؤ لـ  $Y$ . أو  
 بعبارة أخرى هل هناك تداخل معنوي بين  $x_1, x_2$ ؟

**الحل**

- (أ) تحت فرض نموذج الانحدار (٦-١١) فإن جدول تحليل التباين معطى في جدول (٦-٢٦).

جدول (٦-٢٦)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12}   \beta_0$	5	365.477	73.095	174.179
الخطأ	12	5.036	.420	
الكلي	17	370.512		

بما أن قيمة  $F$  المحسوبة من جدول (٦-٢٦) تزيد عن قيمة  $F$  الجدولية  $F_{0.05}(5,12) = 3.11$  فإننا نرفض فرض العدم ونقبل للفرض البديل. أي أن هناك على الأقل إحدى معاملات الانحدار الجزئية لا تساوي صفراً.

- (ب) لاختبار هل هناك تداخلا بين  $x_1, x_2$  فإن فرض العدم سيكون:

$$H_0 : \beta_{12} = 0.$$

ضد الفرض البديل:

$$H_1 : \beta_{12} \neq 0.$$

جدول تحليل التباين عندما تكون الحدود  $x_1, x_2, x_1^2, x_2^2$  موجودة في النموذج معطى في جدول (٦-٢٧).

جدول (٦-٢٧)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2, \beta_{11}, \beta_{22}   \beta_0$	4	364.244	91.061	188.853
الخطأ	13	6.268	.482	
الكل	17	370.512		

من جدول (٦-٢٦) وجدول (٦-٢٧) فلنأخذ نصيب:

$$\begin{aligned}
 SSR(\beta_{12} | \beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_0) &= SSR(\beta_1, \beta_2, \beta_{11}, \beta_{12}, \beta_{22} | \beta_0) \\
 &- SSR(\beta_1, \beta_2, \beta_{11}, \beta_{22} | \beta_0) \\
 &= 365.477 - 364.244 \\
 &= 1.233.
 \end{aligned}$$

قيمة F نصيب كالتالي:

$$\begin{aligned}
 F &= \frac{MSR(\beta_{12} | \beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_0)}{MSE(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12} | \beta_0)} \\
 &= \frac{1.233}{0.42} = 2.9357.
 \end{aligned}$$

وبما أن قيمة F المحسوبة أقل من القيمة الجدولية  $F_{0.05}(1,12) = 4.75$  فلنأخذ  
نقبل فرض العدم، أي أنه لا يوجد تداخل بين المتغيرين  $X_1, X_2$ .

#### تحديد درجة المعادلة

١- الطريقة العكسية: لتحديد درجة المعادلة التي تحتوي على أكثر من  
متغير واحد يستحسن استخدام الطريقة العكسية في ذلك والتي سوف نتناولها  
بالمثال التالي:



مثال (١١-٦)

بالرجوع إلى مثالنا (١٠-٦) فإننا قد أوجدنا معادلة الحدار من الدرجة الثانية. والآن نبدأ بتحديد درجة المعادلة كالتالي:

(١) نختبر هل حدود الدرجة الثانية تساعد معنويا على التنبؤ بـ  $Y$ . أي اختبار فرض العدم:

$$H_0: \beta_{11} = 0, \beta_{12} = 0, \beta_{22} = 0$$

لعمل هذا الاختبار من جدول (٢٦-٦) و جدول (٢٨-٦) نحسب:

$$\begin{aligned} & SSR(\beta_{11}, \beta_{22}, \beta_{12} | \beta_1, \beta_2, \beta_0) \\ &= SSR(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12} | \beta_0) \\ &- SSR(\beta_1, \beta_2, \beta_0) \\ &= 365.477 - 346.510 \\ &= 18.967. \end{aligned}$$

جدول (٢٨-٦)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2, \beta_0$	2	346.510	173.272	108.272
الخطأ	15	24.003	1.600	
الكل	17	370.512		

قيمة F تحسب كالتالي:

$$\begin{aligned} F &= \frac{SSR(\beta_{11}, \beta_{22}, \beta_{12} | \beta_1, \beta_2, \beta_0) / 3}{MSE(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12} | \beta_0)} \\ &= \frac{(18.967) / 3}{0.420} = 15.053. \end{aligned}$$

وبما أن قيمة  $F$  المحسوبة تزيد عن قيمة  $F$  الجدولية  $F_{0.05}(3,12) = 3.49$  لذا فإنه ليست كل حدود الدرجة الثانية تساوي صفراً.  
الآن نختبر كل حد من حدود الدرجة الثانية على حده. وذلك باستخدام جدول (٢٦-٦) و جدول (٢٩-٦) في حساب:

$$\begin{aligned} & SSR(\beta_{11}|\beta_1, \beta_2, \beta_{22}, \beta_{12}, \beta_0) \\ &= SSR(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12}|\beta_0) - SSR(\beta_1, \beta_2, \beta_{22}, \beta_{12}|\beta_0) \\ &= 365.477 - 364.443 \\ &= 1.034. \end{aligned}$$

#### جدول (٢٩-٦)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2, \beta_{22}, \beta_{12} \beta_0$	4	364.443	91.111	195.147
الخطأ	13	6.069	.467	
الكل	17	370.512		

قيمة  $F$  تحسب كالتالي:

$$\begin{aligned} F &= \frac{MSR(\beta_{11}|\beta_1, \beta_2, \beta_{22}, \beta_{12}, \beta_0)}{MSE(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12}|\beta_0)} \\ &= \frac{1.034}{0.42} = 2.4619. \end{aligned}$$

وبما أن قيمة  $F$  المحسوبة تقل عن قيمة  $F$  الجدوليه  $F_{0.05}(1,12) = 4.75$  فإننا نقبل فرض العدم أي أن  $\beta_{11} = 0$ . الآن من جدول (٢٦-٦) و جدول (٣٠-٦) نحسب:

$$\begin{aligned}
 & SSR(\beta_{22}|\beta_1, \beta_2, \beta_{11}, \beta_{12}, \beta_0) \\
 &= SSR(\beta_1, \beta_2, \beta_{11}, \beta_{22}, \beta_{12}|\beta_0) \\
 &- SSR(\beta_1, \beta_2, \beta_{11}, \beta_{12}|\beta_0) \\
 &= 365.477 - 348.776 \\
 &= 16.701.
 \end{aligned}$$

تحسب قيمة F من الصيغة التالية:

$$F = \frac{16.701/1}{0.42} = 39.76.$$

و بما أن قيمة F المحسوبة تزيد عن قيمة F الجدولية  $F_{0.05}(1,12) = 4.75$  فإننا نرفض فرض العدم. أي أن  $\beta_{22} \neq 0$ . وأخيرا أثبتنا من قبل أن  $\beta_{12} = 0$  وعلى ذلك فإننا نحذف الحدين  $x_1^2$ ,  $x_1 x_2$  من المعادلة لعدم أهميتها ونوجد المعادلة التي تحتوي على الحدود  $x_1, x_2, x_2^2$  حيث :

$$\hat{y} = 42.367 - .136x_1 - 2.439x_2 + .08173x_2^2.$$

جدول (٦-٣٠)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2, \beta_{11}, \beta_{12} \beta_0$	4	348.776	87.194	52.148
الخطأ	13	21.737	1.672	
الكلي	17	370.512		

#### طريقة كثيرات الحدود

ويستخدم في التجارب العاملية لتحديد درجة المعادلة وتقدير معالمها وتستخدم عندما تكون جميع العوامل الداخلة في التجارب العاملية كمية وذات مستويات متساوية المسافات ويمكن الرجوع إلى كتاب تصميم التجارب وتحليلها للمؤلفة لتناول هذا الجزء بالتفصيل في الفصل الخامس الخاص بالتجارب السلية.

## الفصل السابع المتغيرات الصورية Dummy Variables

- (١-٧) المتغيرات الصورية في حالة متغيرات مستقلة وصفية
- (٢-٧) متغير مستقل وصفي بمستويين
- (٣-٧) متغير مستقل وصفي بأكثر من مستويين
- (٤-٧) حالة أكثر من متغير صوري في نموذج الانحدار
- (٥-٧) تطبيقات المتغيرات الصورية في السلاسل الزمنية
- (٦-٧) نماذج الانحدار بمتغيرات صورية تخص متغير الاستجابة
- (١-٦-٧) النموذج الخطي
- (٢-٦-٧) النموذج الغير خطي

## (٧-١) المتغيرات الصورية في حالة متغيرات مستقلة وصفية

يهتم تحليل الانحدار في معظم الحالات بالمتغيرات المستقلة الكمية مثل الانتاج ودرجة الحرارة ، الدخل ، المسافة ، الضغط وغيرها من المتغيرات الكمية ، اي المتغيرات التي تقاس وتأخذ قيما معينة. في كثير من مجالات الاعمال ، الاقتصاد ، العلوم الاجتماعية ، الحيوية لا تكون دائما المتغيرات المستقلة كمية ولكن دائما وصفية (متغيرات المجاميع) ، مثل الجنس (ذكر ، انثى) ، الحالة السياسية (حرب أو سلام) ، وجود المرضى (نعم أو لا) ، التدخين (يدخن أو لايدخن) ، فصول السنة (شتاء ، ربيع ، خريف ، صيف) ، الحالة الاجتماعية (ارمل - مطلق - اعزب - متزوج).

سوف نعرف المتغيرات المستقلة الوصفية في نموذج الانحدار بطريقة كمية وذلك باستخدام متغيرات صورية dummy variable (أو المتغيرات المعيرة أو المؤشر indicator variables) والتي تأخذ قيما : الواحد أو الصفر. وكقاعدة عامة إذا كان المتغير المستقل له  $k$  من المستويات أو الفئات أو الأرقام أو المجاميع فإنه يمكن تمثيله بـ  $k-1$  من المتغيرات الصورية.

إما إذا قمنا بتعريف  $k$  متغيرا صوريا بعدد مستويات المتغير المستقل في حالة اشتغال نموذج الانحدار على المعامل الثابت ( $\beta_0$ ) فإننا نواجه مشكلة الارتباط الخطي التام ، أي المشكلة الذي يتعذر بوجودها استخدام طريقة المربعات الصغرى لتقدير معالم النموذج والتي سوف نتناولها بالتفصيل في الفصل التاسع. لمثلا إذا كان لدينا متغير صوري وله فئتين كمتغير الجنس وقمنا بتعريف متغيرين صوريين لتمثيل صفتي المتغير، نجد أن المصفوفة  $X$  تأخذ الشكل التالي:

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & 0 & 1 \\ 1 & \vdots & \vdots \\ 1 & 0 & 1 \end{bmatrix}$$

حيث يحتوي العمود الأول على القيمة واحد لتقدير المعامل الثابت والعمودين الثاني والثالث يحتويان على قيم المتغيرين الصوريين. ويلاحظ من المصفوفة  $X$

انه يمكن الحصول على قيم العمود الثاني بطرح قيم العمود الثالث من قيم العمود الاول وكذلك يمكن الحصول على قيم العمود الثالث من المصفوفة  $X$  بطرح قيم العمود الثاني من قيم العمود الاول. وبالتالي نجد أن محدد المصفوفة يساوى صفر ومن ثم لا يمكن إيجاد معكوس المصفوفة  $X'X$  وعدم امكانية استخدام طريقة المربعات الصغرى لتقدير معالم النموذج. ويلاحظ عدم بروز هذه المشكلة في حالة عدم احتواء نموذج الانحدار على المعامل الثابت  $\beta_0$  حيث تأخذ المصفوفة  $X$  في هذا المثال الشكل التالي:

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}$$

أمثله

١-جنس عائل الأسرة قد يكون ذكرا أو انثى وعليه فإن  $k=2$  في هذه الحالة وعدد المتغيرات الصورية التي تمثل الجنس سوف تكون  $k-1=1$ . فإذا رمزنا للمتغير (الجنس) بالرمز  $x_1$  فيمكن لهذا المتغير اخذ القيم 1 أو 0 كما يلي :

$$x_1 = \begin{cases} 1 & \text{إذا كان عائل الأسرة أنثى} \\ 0 & \text{إذا كان عائل الأسرة ذكر} \end{cases}$$

فإذا كان لدينا أربعة أسر وكان عائل الأسرة فيها كالاتي:

رقم الأسرة	$x_1$
1	M
2	M
3	F
4	M

حيث  $M$  ترمز لذكر و  $F$  ترمز لأنثى فإن المتغير  $x_1$  يأخذ القيم التالية:

رقم الأسرة	$x_1$
1	0
2	0
3	1
4	0

عادة تختار المجموعة الأولى (التي تأخذ القيمة 1) عشوائيا. تسمى المجموعة التي تأخذ القيمة صفر بمجموعة الأساس أو المرجع reference. أن القسم المعنية (0 و 1) للمتغير الصوري ليس هدفها إعطاء الأهمية الكمية لمجاميع المتغير الوصفي بل أن هدفها هو التمييز بين المجاميع. ويمكن استخدام متغيرات أخرى بدلا من المتغيرات الصورية. ففي بعض التطبيقات قد تعطى أرقاما غير الصفر والواحد لفئات المتغير المستقل الوصفي . فعلى سبيل المثال عند دراسة العلاقة بين الراتب السنوي وعدد سنوات الخدمة  $x_1$  والحالة التعليمية  $x_2$  حيث  $x_2$  متغير مستقل وصفي وله ثلاثة فئات (حاصل على شهادة الثانوية - حاصل على مؤهل عالي - حاصل على شهادة عليا) والتي تعطى الأرقام 3, 2, 1 على التوالي . أن القيم 3, 2, 1 قيم عشوائية وقد تعطى أرقاما أخرى . وقد وجد بأن الواحد والصفر هي أكثر الأرقام استعمالا من غيرها وبإسقاطها في تفسير النتائج .

في دراستنا لموضوع المتغيرات المستقلة الوصفية سوف نكتفي بالمتغيرات الصورية أي التي تأخذ القيمة واحد لمستوي من مستويات المتغير المستقل والقيمة صفر لبقية المستويات .

٢- عند دراسة تأثير كل من عدد العاملين في مطعم وموقع المطعم على المبيعات وذلك لمسلسلة من المطاعم فقد يقسم موقع المطعم إلى:

- طريق سريع
- مجمع تجارى
- شارع

ففي هذه الحالة  $k = 3$  وإذلك كان عدد المتغيرات الصورية التي يمكن استخدامها لهذا المتغير هي  $k-1=3-1=2$ . سوف نرمز لتلك المتغيرات الصورية بالرمزين  $x_2, x_3$  ونعرفهم كالآتي:

$$x_2 = \begin{cases} 1 & \text{إذا كان الموقع مجمع تجاري} \\ 0 & \text{غير ذلك} \end{cases}$$

و

$$x_3 = \begin{cases} 1 & \text{إذا كان الموقع شارع} \\ 0 & \text{غير ذلك} \end{cases}$$

وعلى ذلك القيم العددية  $x_2, x_3$  والتي ترتبط بتلك المواقع الثلاثة سوف تكون:

الموقع	$x_2$	$x_3$
الطريق السريع	0	0
مجمع تجاري	1	0
الشارع	0	1

#### (٧-٢) متغير مستقل وصفي بمستويين

لا توجد مشاكل لعمليات حسابيه جديدة عندما يمثل متغير مستقل في نموذج الانحدار بغئة من المتغيرات الصورية. العنصر الوحيد الجديد هو تفسير معاملات الانحدار للمتغيرات الصورية والمثال التالي سوف يوضح ذلك.

#### مثال (٧-١)

تبيع شركة للأجهزة المكتبية حاسبات يدوية مستوردة بموجب امتياز وتقوم بصيانة وقائية وخدمة اصلاح لتلك الحاسبات. البيانات المعطاه في جدول (٧-١) لـ 18 طلبا حديثا من مستخدمي الحاسبات للقيام بخدمة وصيانة وقائية روتينية ، ولكل طلب يمثل  $x_1$  عدد الحاسبات التي تتطلب صيانة ونوع الآلة الحاسبة  $x_2$  و  $y$  عدد الدقائق التي يستغرقها اداء خدمة مطلوبة .



جدول (١-٧)

5	8	2	4	3	5	1	5	7	$x_1$
C	C	C	C	C	C	C	C	C	$x_2$
71	118	33	53	39	75	10	78	97	y
7	2	5	7	4	6	5	4	1	$x_1$
S	S	S	S	S	S	C	C	C	$x_2$
105	25	65	101	62	86	68	49	17	y

من جدول (١-٧) يتضح أن المتغير المستقل  $x_1$  (عدد الحاسبات التي تتطلب صيانة) هو متغير كمي بينما المتغير المستقل  $x_2$  (نوع الحاسب) هو متغير وصفي وله مستويين S و C وبما أن المتغير الوصفي  $x_2$  له مستويان فيمكن تمثيله بمتغير صوري واحد فقط ( $k=1$ ) وهو:

$$x_2 = \begin{cases} 1 & \text{نوع الحاسب S} \\ 0 & \text{نوع الحاسب C} \end{cases}$$

نموذج الانحدار سوف يكون:

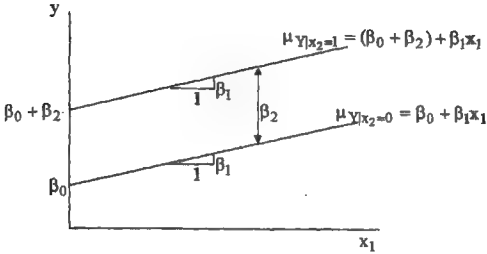
$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon. \quad (١-٧)$$

النموذج (١-٧) يحتوي على متغير كمي والآخر وصفي ولذلك يسمى نموذج تحليل التباين المشترك والذي يختلف عن نموذج الانحدار الذي يحتوي على متغيرات مستقلة كلها وصفية ويسمى نموذج تحليل التباين والنموذج الأخير يمكن الرجوع له بالتفصيل في كتاب تصميم وتحليل التجارب الخاص بالمولفة. توفيق النموذج (١-٧) يكافئ توفيق نموذجين انحدار منفصلين. لتفسير المعامل في النموذج (١-٧) وبفرض نوع الحاسب C حيث  $x_2 = 0$  فإن دالة الاستجابة سوف تكون:

$$\mu_{Y|x_1, x_2} = \beta_0 + \beta_1 x_1 + \beta_2 (0) = \beta_0 + \beta_1 x_1.$$

وعلى ذلك إذا كان الحاسب من نوع C فإن العلاقة بين  $x_1$  (عدد الآلات المخدومة وعدد الدقائق التي يستغرقها أداء خدمة مطلوبة Y عبارة عن خط

مستقيم بمعامل انحدار يساوى  $\beta_1$  ونقطة تقاطع  $\beta_0$  كما هو موضح في شكل (١-٧).



شكل (١-٧)

وبفرض نوع الحاسب S أي  $x_2 = 1$  فإن دالة الاستجابة ستكون:

$$\begin{aligned}\mu_{Y|x_1, x_2} &= \beta_0 + \beta_1 x_1 + \beta_2(1) \\ &= (\beta_0 + \beta_2) + \beta_1 x_1\end{aligned}$$

أي انه للنوع S فإن العلاقة بين  $x_1$  و  $Y$  تمثل بخط مستقيم بمعامل انحدار  $\beta_1$  أيضا يساوى  $\beta_1$  ولكن بنقطة تقاطع هي  $\beta_0 + \beta_2$  كما هو موضح في شكل (١-٧). لذا فإن  $\beta_2$  هي مقدار ارتفاع أو انخفاض دالة الاستجابة من الفئة (الواحد) عن الخط للفئة (صفر). وبالعودة إلى مثالنا (١-٧) واعطاء 1 للحاسب من نوع S و 0 للحاسب من نوع C فإن المصفوفة X والمتجه y انموذج الانحدار يكونان:

$$X = \begin{bmatrix} 1 & 7 & 0 \\ 1 & 5 & 0 \\ 1 & 1 & 0 \\ 1 & 5 & 0 \\ 1 & 3 & 0 \\ 1 & 4 & 0 \\ 1 & 2 & 0 \\ 1 & 8 & 0 \\ 1 & 5 & 0 \\ 1 & 1 & 0 \\ 1 & 4 & 0 \\ 1 & 5 & 0 \\ 1 & 6 & 1 \\ 1 & 4 & 1 \\ 1 & 7 & 1 \\ 1 & 5 & 1 \\ 1 & 2 & 1 \\ 1 & 7 & 1 \end{bmatrix}, y = \begin{bmatrix} 97 \\ 78 \\ 10 \\ 75 \\ 39 \\ 53 \\ 33 \\ 118 \\ 71 \\ 17 \\ 49 \\ 68 \\ 86 \\ 62 \\ 101 \\ 65 \\ 25 \\ 105 \end{bmatrix}$$

معادلة الانحدار المقدرة سوف تكون :

$$\hat{y} = -2.348 + 14.723 x_1 + 0.277 x_2.$$

جدول تحليل التباين معطى في جدول (٧-٢).

جدول (٧-٢)

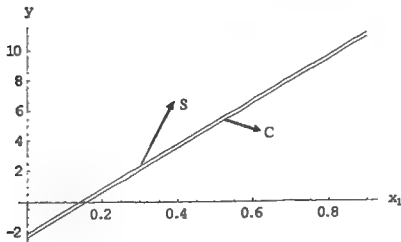
S.O.V	df	SS	MS	F
$\beta_1, \beta_2 / \beta_0$	2	16182.894	8091.447	377.980
الخطأ	15	321.106	21.407	
الكلي	17	16504.000		

ايضا فهم t مع المعنوية المقابلة لها معطاء في جدول (٧-٣) .

جدول (٧-٣)

المعامل	التقدير	الخطا المعياري	t	p-value المعنوية
$\beta_0$	-2.348	2.656	-0.884	0.391
$\beta_1$	14.723	0.551	26.719	0.000
$\beta_2$	0.277	2.378	0.116	0.909

ومن جدول (٧-٢) وبما أن قيمة F المحسوبة تزيد عن القيمة الجدولية  $F_{0.05}(2,15) = 3.68$  فهذا يعني أن الانحدار معنوي. وبما أن قيمة  $|t|$  في جدول (٧-٣) الخاصة بمعلمة  $\beta_1$  معنوية فهذا يعني أن  $\beta_1 \neq 0$  أي أن المتغير المستقل  $x_1$  يساعد معنوياً على التنبؤ بـ  $Y$ . إن عدم معنوية  $\beta_2$  تدل على أن خطي الانحدار في شكل (٧-٢) للحاسب من نوع S والحاسب من نوع C متطابقين. أي أن نقطتي تقاطع خطي الانحدار لا تختلفان معنوياً. وبالتالي يكون لدينا خط انحدار واحد. إن معنوية  $\beta_1$  تعني أن زيادة عدد الحاسبات المخدومة بمقدار واحد سيزيد عدد الدقائق التي يستغرقها في الصيانة بمقدار  $\beta_1$ . أما  $\beta_2$  فهي التغير في عدد الدقائق التي يستغرقها في الصيانة للحاسب من نوع S إلى الحاسب من نوع C وذلك عندما تكون  $\beta_2$  معنوية.



شكل (٧-٢)

### التفاعل بين المتغيرات الوصفية والكمية

اوضحنا فيما سبق كيفية تأثير المتغير الوصفي على المعامل الثابت ولكن لم ندرس أثر المتغير الوصفي على ميل الانحدار. ولقياس اثر المتغير الوصفي على الميل يتم إضافة متغير مستقل مركب عبارة عن مضروب المتغير الصوري في المتغير الكمي ويعرف هذا المتغير بمتغير التفاعل حيث يقيس الأثر المشترك للمتغيرين الوصفي والكمي على المتغير التابع. ولتوضيح اثر المتغير الوصفي على ميل نموذج الانحدار سوف نفترض المثال (٧-١) النموذج التالي:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon \quad (٧-٢)$$

وبمقارنة (٧-٢) مع (٧-١) نلاحظ أن حاصل الضرب بين  $x_1$  و  $x_2$  قد اضيفت الى النموذج. لتفسير معالم هذا النموذج ، وبفرض أنه للحاسب من نوع C حيث  $x_2 = 0$  فإن النموذج (٧-٢) يصبح :

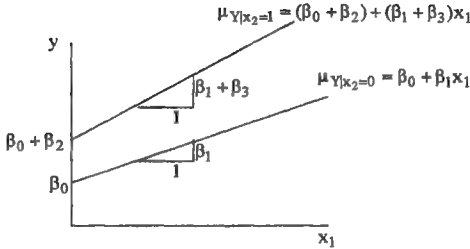
$$\begin{aligned} \mu_{Y|x_1, x_2} &= \beta_0 + \beta_1 x_1 + \beta_2 (0) + \beta_3 x_1 (0) \\ &= \beta_0 + \beta_1 x_1 \end{aligned}$$

أي أن دالة الاستجابة للحاسب من نوع C عبارة عن خط مستقيم بمعامل انحدار  $\beta_1$  ونقطة تقاطع  $\beta_0$  كما هو موضح في شكل (٧-٣).

ايضا للحاسب من نوع S حيث  $x_2 = 1$  فإن دالة الاستجابة سوف تكون:

$$\begin{aligned} \mu_{Y|x_1, x_2} &= \beta_0 + \beta_1 x_1 + \beta_2 (1) + \beta_3 x_1 (1) \\ &= (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_1 \end{aligned}$$

أي أن دالة الاستجابة في حالة الحاسب من نوع S هي خط مستقيم ايضا ولكن بمعامل انحدار  $\beta_1 + \beta_3$  ونقطة تقاطع  $\beta_0 + \beta_2$  .



شكل (٣-٧)

كلا الخطين موضحين في شكل (٣-٧). ويجب ان نتذكر ان (٣-٧) تعرف خطين انحدار بميلين مختلفين ونقطتي تقاطع مختلفتين. وعلى ذلك المعطاة  $\beta_2$  تعكس التغير (بالزيادة أو النقصان) في الجزء المقطوع المرتبط بالتغير من الحاسب من نوع S الى الحاسب من نوع C و  $\beta_3$  توضح التغير في الميل المرتبط بالتغير من الحاسب S الى الحاسب C. توفيق النموذج (٢-٧) يكافئ توفيق نموذجين انحدار منفصلين. من مزايا استخدام المتغيرات الصورية ان اختبارات الفروض يمكن اجرائها مباشرة باستخدام طريقة مجاميع المربعات الإضافية. على سبيل المثال لاختبار فيما إذا كان النموذجين متطابقين، فإننا نختبر فرض العدم:

$$H_0 : \beta_2 = \beta_3 = 0,$$

ضد الفرض البديل :

$$H_1 : \beta_2 \neq 0 \text{ (أو) } \beta_3 \neq 0$$

عند قبول فرض العدم  $H_0 : \beta_2 = \beta_3 = 0$  فهذا يعني أن نموذج الانحدار واحد كان يكفي لشرح العلاقة بين  $Y$ ,  $x_1$  . لاختبار أن خطين الانحدار لهما ميل واحد ولكن من الممكن مختلفتين في الجزء المقطوع من محور  $Y$  فإننا نختبر فرض العدم:

$$H_0 : \beta_3 = 0$$

ضد الفرض البديل :

$$H_1 : \beta_3 \neq 0$$

للمثال (١-٧) سوف نقوم بإيجاد معادلة الانحدار المقترنة لنموذج الانحدار :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon .$$

لتوفيق البيانات المعطاه في جدول (١-٧) فإن المصفوفة  $X$  والمتجه  $y$  لهذا النموذج هما:

$$X = \begin{bmatrix} x_1 & x_2 & x_1x_2 \\ 1 & 7 & 0 & 0 \\ 1 & 5 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 5 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 8 & 0 & 0 \\ 1 & 5 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 1 & 5 & 0 & 0 \\ 1 & 6 & 1 & 6 \\ 1 & 4 & 1 & 4 \\ 1 & 7 & 1 & 7 \\ 1 & 5 & 1 & 5 \\ 1 & 2 & 1 & 2 \\ 1 & 7 & 1 & 7 \end{bmatrix}, y = \begin{bmatrix} 97 \\ 78 \\ 10 \\ 75 \\ 39 \\ 53 \\ 33 \\ 118 \\ 71 \\ 17 \\ 49 \\ 68 \\ 86 \\ 62 \\ 101 \\ 65 \\ 25 \\ 105 \end{bmatrix}$$

نموذج الانحدار المقدر سيكون :

$$\hat{y} = -1.565 + 14.535x_1 - 3.170x_2 + 0.703x_1x_2 .$$

جدول تحليل التباين عندما  $x_1, x_2, x_1x_2$  موجودين في نموذج الانحدار معطى في جدول (٤-٧) .

جدول (٤-٧)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2, \beta_3   \beta_0$	3	16189.724	5396.575	240.400
الخطأ	14	314.276	22.448	
الكلية	17	16504.000		



جدول تحليل التباين عندما  $x_1$  موجود فقط في نموذج الانحدار معطى في جدول (٥-٧) .

جدول (٥-٧)

S.O.V	df	SS	MS	F
$\beta_1   \beta_0$	1	16182.604	16182.604	805.616
الخطأ	16	321.396	20.087	
الكلية	17	16504.000		

جدول تحليل التباين عندما  $x_1, x_2$  موجودين فقط في نموذج الانحدار معطى في جدول (٦-٧) .

جدول (٦-٧)

S.O.V	df	SS	MS	F
$\beta_1, \beta_2   \beta_0$	2	16182.894	8091.447	377.980
الخطأ	15	321.106	21.407	
الكلية	17	16504.000		

لاختبار فرض العدم ان خطين الانحدار متطابقين نختبر  $H_0: \beta_2 = \beta_3 = 0$  وذلك باستخدام الإحصاء :

$$F = \frac{SSR(\beta_2, \beta_3 | \beta_1, \beta_0) / 2}{MSE(\beta_1, \beta_2, \beta_3 | \beta_0)}$$

ومن جدول (٤-٧) و (٥-٧) فإن :

$$\begin{aligned} SSR(\beta_2, \beta_3 | \beta_1, \beta_0) &= SSR(\beta_1, \beta_2, \beta_3 | \beta_0) \\ &\quad - SSR(\beta_1 | \beta_0) \\ &= 16189.724 - 16182.604 \\ &= 7.12. \end{aligned}$$

قيمة F المحسوبة سوف تكون :

$$F = \frac{SSR(\beta_2, \beta_3 | \beta_1, \beta_0) / 2}{MSE(\beta_1, \beta_2, \beta_3 | \beta_0)} = \frac{3.56}{22.448} = 0.1586 .$$

وبما أن قيمة  $F$  المحسوبة أقل من القيمة الجدولية  $F_{0.05}(3,14) = 3.34$  فإننا نستنتج أن الخططين متطابقين كما يتضح من شكل (٧-٤) . لاختبار فرض العدم أن الخططين ربما لهما جزء مقطوع مختلف وميل واحد ( $H_0 : \beta_3 = 0$ ) يستخدم الإحصاء :

$$F = \frac{SSR(\beta_3 | \beta_1, \beta_2, \beta_0) / 1}{MSE(\beta_1, \beta_2, \beta_3 | \beta_0)} .$$

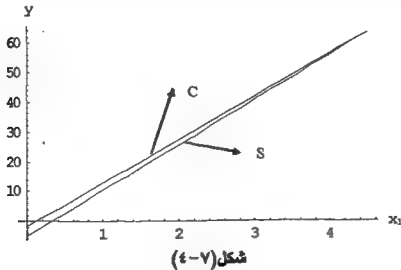
ومن جدول (٧-٤) فإن :

$$\begin{aligned} SSR(\beta_3 | \beta_1, \beta_2, \beta_0) &= SSR(\beta_1, \beta_2, \beta_3 | \beta_0) - SSR(\beta_1, \beta_2 | \beta_0) \\ &= 16189.724 - 16182.894 \\ &= 6.83 . \end{aligned}$$

قيمة  $F$  المحسوبة سوف تكون :

$$F = \frac{SSR(\beta_3 | \beta_1, \beta_2, \beta_0) / 1}{MSE(\beta_1, \beta_2, \beta_3 | \beta_0)} = \frac{6.83}{22.448} = 0.3043 .$$

وبما أن قيمة  $F$  المحسوبة أقل من القيمة الجدولية  $F_{0.05}(1,14) = 4.6$  فإننا نستنتج أن الميل للخططين واحد . أيضا يمكن استخدام اختبار  $t$  لكل من  $\beta_3, \beta_2, \beta_1$  .



شكل (٧-٤)

يعطي جدول (٧-٧) قيم  $t$  مع قيم المعنوية الخاصة بها .

جدول (٧-٧)

المعامل	التقدير	الخطأ المعياري	t	p-value المعنوية
$\beta_0$	-1.565	3.068	-0.510	0.618
$\beta_1$	14.535	0.659	22.052	0.000
$\beta_2$	-3.170	6.706	-0.473	0.644
$\beta_3$	0.703	1.275	0.552	0.590

يتضح من جدول (٧-٧) معنوية  $\beta_1$  فقط .

(٣-٧) متغير مستقل وصفي بأكثر من مستويين

مثال (٧-٢)

اجريت دراسة على سلسلة من المطاعم لمعرفة العلاقة بين مبيعات المطعم خلال فترة من الزمن (Y بالآلاف دولار) وعدد العاملين في المطعم ( $x_1$ ) وموقع المطعم (Street -Mall- Highway) . المستويات الثلاثة لعامل الموقع يمكن تمثيلها بمتغيرين صوريين  $x_2, x_3$  يعرفان كالآتي :

الموقع	$x_2$	$x_3$
Highway طريق سريع	0	0
Mall مجمع تجارى	1	0
Street شارع	0	1

نموذج الاتحاد سوف يكون:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

البيانات اللازمة لتفريق هذا النموذج معطاه في جدول (٨-٧).

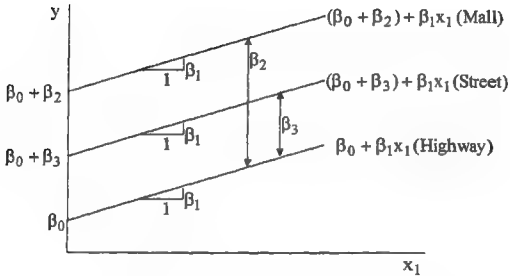
جدول (٨-٧)

$x_1$	الفئة الوصفية	$x_2$	$x_3$	$y$
155	Highway	0	0	135.27
93	Highway	0	0	72.74
128	Highway	0	0	114.95
114	Highway	0	0	102.93
158	Highway	0	0	131.77
183	Highway	0	0	160.91
178	Mall	1	0	179.86
215	Mall	1	0	220.14
172	Mall	1	0	179.64
197	Mall	1	0	185.92
207	Mall	1	0	207.82
95	Mall	1	0	113.51
224	Street	0	1	203.98
199	Street	0	1	174.48
240	Street	0	1	220.43
100	Street	0	1	93.19

ولكي نفهم معنى معاملات الاتحاد لهذا النموذج وبفرض مطعم في Highway حيث  $x_2 = 0$ ,  $x_3 = 0$  فإن دالة الاستجابة  $\mu_{Y|x_1, x_2, x_3}$  سوف تختزل إلى الشكل التالي:

$$\begin{aligned} \mu_{Y|x_1, x_2, x_3} &= \beta_0 + \beta_1 x_1 + \beta_2 (0) + \beta_3 (0) \\ &= \beta_0 + \beta_1 x_1 \end{aligned}$$

حيث المعامل الثابت  $\beta_0$  يمثل نقطة تقاطع خط انحدار فئة الأساس (مطعم Highway). أي أن  $\mu_{Y|x_1, x_2, x_3}$  خط مستقيم بجزء مقطوع من المحور Y يساوي  $\beta_0$  وميل يساوي  $\beta_1$  ودالة الاستجابة موضحة في شكل (٥-٧).



شكل (٥-٧)

لمطعم في Mall حيث  $x_3 = 0, x_2 = 1$  فإن دالة الاستجابة تصبح على الشكل التالي:

$$\begin{aligned}\mu_{Y|x_1, x_2, x_3} &= \beta_0 + \beta_1 x_1 + \beta_2 (1) + \beta_3 (0) \\ &= (\beta_0 + \beta_2) + \beta_1 x_1\end{aligned}$$

وهذا أيضا خط مستقيم بميل يساوي  $\beta_1$  ولكن الجزء المقطوع مع محور Y هو  $(\beta_0 + \beta_2)$  حيث  $\beta_2$  تمثل الفرق في نقطة التقاطع بين خط انحدار فئة الأساس (مطعم في Highway) ومطعم في Mall ، ودالة الاستجابة موضحة أيضا في شكل (٥-٧). وأخيرا لمطعم في Street عندما  $x_3 = 1, x_2 = 0$  فإن دالة الاستجابة تصبح على الشكل :

$$\begin{aligned}\mu_{Y|x_1, x_2, x_3} &= \beta_0 + \beta_1 x_1 + \beta_2 (0) + \beta_3 (1) \\ &= (\beta_0 + \beta_3) + \beta_1 x_1\end{aligned}$$

والذي يمثل خط مستقيم بجزء مقطوع مع محور  $Y$  يساوي  $(\beta_0 + \beta_3)$  ويميل يساوي  $\beta_1$  كما هو موضح في شكل (٥-٧) . وبسبب ان الخطوط الثلاثة متوازية فإنه لأي قيمة عددية معطاء للمتغير المستقل  $x_1$  فإن متوسط الاستجابة لمطعم في Mall تختلف عن مطعم في Highway بمقدار  $\beta_2$  ومتوسط الاستجابة لمطعم في Street يختلف عن مطعم في Highway بمقدار  $\beta_3$  . شكل (٥-٧) يوضح كيف أن  $\beta_2, \beta_3$  يعكس تأثير الاختلاف في موقع Street و Mall بالنسبة لموقع في Highway ويلاحظ أن دالة الاستجابة لمطعم في Mall تختلف عن مطعم في Street بمقدار  $\beta_2 - \beta_3$  وذلك لأي قيمة معطاء من  $x_1$  . من شكل (٥-٧) يتضح أن  $\beta_1, \beta_2, \beta_3$  موجبتين حيث  $\beta_2$  اكبر من  $\beta_3$  .

معادلة الانحدار المقدرة سوف تكون:

$$\hat{y} = -1.817 + 0.878x_1 + 27.298x_2 + 7.392x_3 .$$

معامل الانحدار  $b_2 = 27.298$  يوضح أنه لأي قيمة رقمية من  $x_1$  فإن متوسط الاستجابة لمطعم في Mall هي 27.3 ألف أكثر من مطعم في Highway . بينما معامل الانحدار  $b_3 = 7.392$  يوضح أنه لأي قيمة رقمية من  $x_1$  فإن متوسط الاستجابة لمطعم في Street هي 7.4 ألف أكثر من مطعم في Highway . وفي النهاية متوسط المبيعات لمطعم في Mall كانت 19.9 ألف دولار أكثر من مطعم في Street حيث:

$$(b_2 - b_3) = 27.298 - 7.392 = 19.906 .$$

وذلك لأي قيمة معطاء من  $x_1$  .

95% فترة ثقة لـ  $\beta_2$  هي:

$$19.410 \leq \beta_2 \leq 35.186 .$$

وعلى ذلك بـ 95% فترة ثقة فإننا نقدر لأي قيمة من  $x_1$  متوسط الاستجابة في موقع Mall بين 19.410 ألف الى 35.186 ألف زيادة عن الموقع Highway . جدول تحليل التباين معطى في جدول (٩-٧) .

جدول (٩-٧)

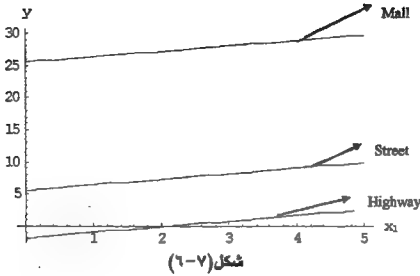
S.O.V	df	SS	MS	F	p-value المعنوية
$\beta_1, \beta_2, \beta_3   \beta_0$	3	33438.857	11146.286	331.403	0.000
الخطأ	12	403.604	33.634		
الكلية	15	33842.460			

من جدول (٩-٧) وبما أن قيمة  $p$  أقل من 0.05. فهذا يعني أن الانحدار معنوي. قيم  $t$  معطاه في جدول (١٠-٧).

جدول (١٠-٧)

المعالم	التقدير	الخطأ المعياري	t	p-value المعنوية
$\beta_0$	-1.817	5.453	-0.333	0.745
$\beta_1$	0.878	0.035	24.752	0.000
$\beta_2$	27.298	3.620	7.54	0.000
$\beta_3$	7.392	4.177	1.77	0.102

من النتائج في جدول (١٠-٧) يتضح معنوية كل من  $\beta_1, \beta_2$  وعدم معنوية  $\beta_3$  عند مستوى معنوية  $\alpha = 0.05$  وذلك لأن  $p = 0.102$  (الخاصة بالمعلمة  $\beta_3$ ) أكبر من قيمة  $\alpha = 0.05$  كما يتضح من شكل (٦-٧) .



#### (٧-٤) حالة أكثر من متغير صوري في نموذج الانحدار

قد يحتوي نموذج الانحدار على أكثر من متغير صوري. بفرض أنه في مثال (٧-١) قد اضيف متغير وصفي ثاني يمثل ميعاد العمل في الشركة (صباحاً M أو مساءً E). سوف نعرف هذا المتغير الصوري الثاني بالرمز  $x_3$  كالآتي:

$$x_3 = \begin{cases} 1 & \text{العمل صباحاً} \\ 0 & \text{العمل مساءً} \end{cases}$$

نموذج الانحدار بدون ادخال متغيرات تفاعل يأخذ الصيغة التالية:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

دوال الاستجابة لهذا التوزيع موضحة في جدول (٧-١١) .



جدول (٧-١١)

$x_2$		$x_3$		دالة الاستجابة
S	1	M	1	$(\beta_0 + \beta_2 + \beta_3) + \beta_1 x_1$
C	0	M	1	$(\beta_0 + \beta_3) + \beta_1 x_1$
S	1	E	0	$(\beta_0 + \beta_2) + \beta_1 x_1$
C	0	E	0	$\beta_0 + \beta_1 x_1$

يتضح من هذا النموذج ان دوال الانحدار المناظرة لصفات المتغيرات الوصفية لها ميل ثابت ونقاط تقاطع مختلفة. وفي حالة إدخال متغيرات تفاعل يأخذ النموذج الصيغة التالية:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3$$

حيث يضم هذا النموذج ثلاثة متغيرات تفاعل. دوال الاستجابة لهذا النموذج معطاه في جدول (٧-١٢).

جدول (٧-١٢)

$x_2$		$x_3$		دالة الاستجابة
S	1	M	1	$(\beta_0 + \beta_2 + \beta_3 + \beta_6) + (\beta_1 + \beta_4 + \beta_5) x_1$
C	0	M	1	$(\beta_0 + \beta_3) + (\beta_1 + \beta_5) x_1$
S	1	E	0	$(\beta_0 + \beta_2) + (\beta_1 + \beta_4) x_1$
C	0	E	0	$\beta_0 + \beta_1 x_1$

وفي هذا النموذج ينصب الاهتمام على إجابة الاسئلة الآتية:

- هل هناك تأثير تفاعل معنوي بين المتغير الصوري الأول والمتغير الكمي  $x_1$ .
- هل هناك تأثير تفاعل معنوي بين المتغير الصوري الثاني والمتغير  $x_1$ .
- هل هناك تأثير تفاعل معنوي بين المتغيرين  $x_2$  ,  $x_3$ .

للجابة على هذه الاسئلة يستخدم اختبار  $F$  الجزئي أو اختبار  $t$  . ولابد من اختبار معنوية متغيرات التفاعل أولا وفي حالة عدم معنويتها يتم اختبار وتفسير المتغيرات الأساسية المكونة لمتغيرات التفاعل. كما يمكن اختبار تساوي دوال الاستجابة الأربعة وذلك باختبار فرض العدم:

$$H_0 : \beta_4 = \beta_5 = 0,$$

مقابل الفرض البديل ، على الأقل واحد من المعامل لا يساوي صفر ، حيث يستخدم الإحصاء  $F$  على الصورة التالية:

$$F = \frac{SSR(\beta_4, \beta_5 | \beta_1, \beta_2, \beta_3, \beta_6, \beta_0) / 2}{MSE} .$$

عندما تزيد قيمة  $F$  المحسوبة عن القيمة الجدولية  $F_{\alpha}(2, n-7)$  نرفض فرض العدم ونقرر أن دوال الاستجابة متوازية. وب نفس الطريقة يمكننا إجراء اختبار تطابق دوال الانحدار وذلك باختبار فرض العدم:

$$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0,$$

في مقابل الفرض البديل أنه على الأقل واحدة من هذه المعامل لا تساوي صفر .

ولإجراء هذا الاختبار يستخدم اختبار  $F$  على الصورة التالية:

$$F = \frac{SSR(\beta_2, \beta_3, \beta_4, \beta_5, \beta_6 | \beta_1, \beta_0) / 5}{MSE} .$$

وإذا زادت قيمة  $F$  المحسوبة على القيمة الجدولية  $F_{\alpha}(5, n-7)$  فإننا نحكم بعدم تطابق دوال الاستجابة. أما إذا قبلنا فرض العدم فهذا يعني أن دوال الاستجابة متطابقة ، اي يمكننا استخدام نموذج الانحدار التالي:

$$Y = \beta_0 + \beta_1 x + \epsilon .$$

#### (٧-٥) تطبيقات المتغيرات الصورية في السلاسل الزمنية

كثيرا من الاقتصاديين يطبقون تحليل الانحدار على السلاسل الزمنية. فعند دراسة انحدار الأرباح ( $Y$ ) على المبيعات  $x$  لفترة زمنية معينة فإن نموذج الانحدار يأخذ الشكل التالي:

$$Y = \beta_0 + \beta_1 x + \epsilon .$$

وإذا كانت الفترة الزمنية تحتوي على فترات مختلفة كأن تكون فصول السنة (الربيع والصيف والخريف والشتاء) أو فترات الحرب وفترات سلام فإنه يمكن استخدام المتغيرات الصورية لإيجاد نموذج رياضي واحد لكل الفترات بدلا من

نموذج رياضي لكل فترة . فعلى سبيل المثال لقياس التغيرات الموسمية في نموذج الانحدار تحت الدراسة يمكن إدخال متغير صوري  $x_2$  ممثل للربع الثاني وبأخذ القيمة 1 إذا ما كانت القيمة ممثلة للربع الثاني وصفر بخلاف ذلك وكذلك يمكن إدخال المتغير  $x_3$  ممثلاً للربع الثالث بنفس الطريقة وبالطبع  $x_4$  ممثلاً للربع الرابع بالإضافة إلى ذلك يمكن إدخال متغير يمثل الاتجاه العام  $t$  على النحو التالي المعطى في جدول (٧-١٣) .

جدول (٧-١٣)

	الربع	$x_2$	$x_3$	$x_4$	$t$
Year 1	1	0	0	0	1
Year 1	2	1	0	0	2
Year 1	3	0	1	0	3
Year 1	4	0	0	1	4
Year 2	1	0	0	0	5
Year 2	2	1	0	0	6
Year 2	3	0	1	0	7
Year 2	4	0	0	1	8

ويكون نموذج الانحدار على الشكل التالي:

$$Y = \beta_0 + \beta_1 t + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 .$$

ومعامل  $t$  هو الاتجاه العام ويقاس التغير الربع سنوي ومعامل  $x_2$  مقدار التغير الناتج من الربع الثاني و  $x_3$  مقدار التغير الناتج من الربع الثالث و  $x_4$  مقدار التغير الناتج من الربع الرابع.

مثال (٧-٣)

يعطى جدول (٧-١٤) بيانات الطاقة المولدة ربع سنوياً بالمليون كيلووات / ساعة بدولة الكويت خلال الفترة (1982-1986).

جدول (٧-١٤)

Year	1stQ	2ndQ	3rdQ	4thQ
1982	1798	3285	4274	2342
1983	2001	3499	4633	2365
1984	2116	3889	5018	2870
1985	2442	4409	5548	3018
1986	2574	4580	6175	3606

ولدراسة الاتجاه العام للربع سنوي والتأثيرات الموسمية لكل ربع سنة على حدة فإن البيانات في جدول (٧-١٤) مع المتغيرات الصورية معطاه في جدول (٧-١٥).

جدول (٧-١٥)

Y	t	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
1798	1	0	0	0
3285	2	1	0	0
4274	3	0	1	0
2342	4	0	0	1
2001	5	0	0	0
3499	6	1	0	0
4633	7	0	1	0
2365	8	0	0	1
2116	9	0	0	0
3889	10	1	0	0
5018	11	0	1	0
2870	12	0	0	1
2442	13	0	0	0
4409	14	1	0	0
5548	15	0	1	0
3018	16	0	0	1
2574	17	0	0	0
4580	18	1	0	0
6175	19	0	1	0
3606	20	0	0	1

معادلة الانحدار المقدره سوف تكون :

$$\hat{y} = 1432.95625 + 83.69375t + 1662.50625x_2 \\ (13.338) \quad (11.262) \quad (13.955) \\ + 2776.0125x_3 + 402.91875x_4 . \\ (23.167) \quad (3.331)$$

القيم بين الاقواس هي قيم t. وقد وجد أن جميع قيم t عاليه ومعنوية وهذا يعني أن جميع معاملات الانحدار عاليه ومعنوية وجميعها لا تساوي الصفر. ويمكن اضافة الجملة التالية بعد المعادلة مباشره حتى يتمكن القارئ من استخدام المعادلة:  
(1982 Q1=1).

#### (٦-٧) نماذج الانحدار بمتغيرات صوريه تخص متغير الاستجابة

في بعض الاحيان فإن متغير الاستجابة في نموذج الانحدار يكون وصفي يفترض له قيمتين فقط. وعلى ذلك متغير الاستجابة يعتبر متغير صوري بقيمة اما 0 او 1 . على سبيل المثال عند دراسة العلاقة بين سرعة صواريخ ارض جو ( $x_1$ ) وإصابة الهدف (الطيارة مثلا) فإن Y يأخذ القيم التالية :

$$Y = \begin{cases} 1 & \text{ان اصاب الصاروخ الهدف} \\ 0 & \text{ان اخطأ الصاروخ الهدف} \end{cases}$$

في هذه الحالة فإن القيمة المتوقعة للاستجابة سوف يكون لها تفسير خاص. بفرض نموذج انحدار بمتغير مستقل واحد حيث:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \dots, n$$

إذا كان  $E(\epsilon_i) = 0$  فإن:

$$\mu_{Y|x_i} = \beta_0 + \beta_1 x_i .$$

وبما أن  $Y|x_i$  يأخذ فقط القيمة 1 أو 0 فإن النموذج الاحتمالي للمتغير  $Y|x_i$  سوف يكون توزيع برنولي. حيث  $P(Y_i = 1) = p_i$  ,  $P(Y_i = 0) = 1 - p_i$  . وبما أن متوسط توزيع برنولي هو  $\mu_{Y|x_i} = p_i$  و

$$E(Y|x_i) = \mu_{Y|x_i} = \beta_0 + \beta_1 x_i = p_i ,$$

فإن متوسط الاستجابة هو احتمال أن  $Y|x_i = 1$  وذلك عندما يأخذ المتغير المستقل القيمة  $x_i$ .

إن توفيق نموذج بمتغير صوري ليس سهلاً. واحد من الصعوبات هو أن تباين الخطأ غير ثابت كما يتضح الآن حيث :

$$\begin{aligned} \text{Var}(\epsilon_i | x_i) &= \text{Var}(Y_i | x_i) = \sigma_{Y|x_i}^2 \\ &= p_i(1 - p_i) \\ &= (\beta_0 + \beta_1 x_i)(1 - \beta_0 - \beta_1 x_i). \end{aligned} \quad (3-7)$$

وذلك لأن تباين توزيع برنولي هو :

$$\text{Var}(Y_i | x_i) = p_i(1 - p_i).$$

المعادلة (3-7) تعني أن تباين الأخطاء غير متجانس وفي الحقيقة تعتمد على قيمة المتغير المستقل  $x_i$ . وهذا يعتبر مخالفة للفروض الأساسية للانحدار. أن استخدام طريقة المربعات المرجحة وذلك بإستخدام اوزان تختار بحيث تتناسب عكسياً مع تباين  $Y|x_i$  سوف يؤدي إلى اختزال هذه المشكلة. أيضاً فإن توزيع الأخطاء لن يكون طبيعياً وذلك لأنه عند كل مستوى ممكن من المتغير المستقل يوجد فقط قيمتين لـ  $\epsilon_i$ . وفي النهاية لأن  $\mu_{Y|x_i}$  هو احتمال أن  $Y|x_i = 1$  عندما قيمة المتغير المستقل تساوي  $x_i$  فإنه من المنطقي أن قيم الاستجابة المتنبأ بها تقع بين 0 و 1، لكل قيم  $x_i$  في المدى للبيانات الأصلية. وفي هذه الحالة لن يكون هناك ضمان لتوفر هذا الشرط. الآن سوف نوضح كيفية إيجاد معادلة الانحدار المقدره وسوف نبدأ بنموذج انحدار خطي ثم نفرض نموذج غير خطي يعتمد على الدالة اللوجستية.

### (١-٦-٧) النموذج الخطي

بفرض أن

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon.$$

حيث  $Y$  متغير صوري يأخذ القيمة 0 أو 1. كما أوضحنا من قبل فإننا سوف نستخدم طريقة المربعات الصغرى المرجحة لتقدير معالم هذا النموذج وذلك لأن تباين الخطأ ليس ثابتاً.

مثال (٤-٧)

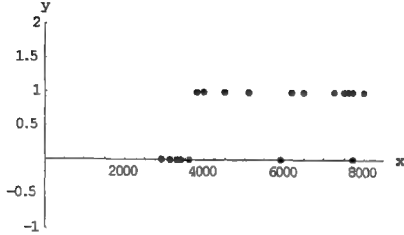
اراد احد الباحثين دراسة العلاقة بين الدخل التصرفي وحالة تملك الدار الساكن فيه لعشرين موظفا والنتائج معطاه في جدول (٧-١٦) وحيث متغير الاستجابة متغير وصفي له فئتين وعليه فمن الممكن تمثيله بمتغير صوري واحد وهو:

$$y = \begin{cases} 1 & \text{الموظف الذي يملك دارا} \\ 0 & \text{الموظف الذي لا يملك دار} \end{cases}$$

جدول (٧-١٦)

x	y	x	y
2900	0	6200	1
7700	1	3400	0
3300	0	6500	1
4500	1	2900	0
5900	0	4000	1
7700	0	8000	1
3800	1	7250	1
3600	0	3100	0
5100	1	3300	0
7500	1	7600	1

شكل الانتشار للبيانات في جدول (٧-١٦) موضح في شكل (٧-٧) .



شكل (٧-٧)

سوف نوجد معادلة الانحدار المقدره للنموذج :

$$Y = \beta_0 + \beta_1 x + \epsilon$$

وذلك باستخدام طريقة المربعات الصغرى المرجحه حيث  $w_i$  هي:

$$w_i = \frac{1}{\sigma_{Y|x_i}^2} = \frac{1}{p_i(1-p_i)}$$

$$= \frac{1}{(\beta_0 + \beta_1 x_i)(1 - \beta_0 - \beta_1 x_i)}$$

و  $w_i$  دالة في معالم مجهولة  $(\beta_0, \beta_1)$ . هذه المشكلة يمكن حلها وذلك بإيجاد معادلة الانحدار المقدره باستخدام طريقة المربعات الصغرى العادية (الغير مرجحه) ثم حساب الاوزان باستخدام تقديرات المربعات الصغرى العادية  $b_0, b_1$  كالتالي:



$$\hat{w}_i = \frac{1}{(b_0 + b_1 x_i)(1 - b_0 - b_1 x_i)}$$

$$= \frac{1}{\hat{y}_i(1 - \hat{y}_i)} .$$

تقديرات المربعات الصغرى العادية لمثالنا (مثال ٧-٤) معطاه في جدول (٧-١٧).

جدول (٧-١٧)

المعامل	التقدير
$\beta_0$	-0.24740598
$\beta_1$	0.000152979

يعطى جدول (٧-١٨) القيم المقدرة  $\hat{y}_i$  والاوزان المقدرة  $\hat{w}_i$  على سبيل المثال عندما  $x_1 = 2900$  فإن :

$$\hat{y}_1 = b_0 + b_1 x_1$$

$$= -0.24740598 + 0.000152979(2900)$$

$$= 0.1962,$$

$$\hat{w}_1 = \frac{1}{\hat{y}_1(1 - \hat{y}_1)} = \frac{1}{(0.1962)(1 - 0.1962)}$$

$$= 6.3401.$$

جدول (٧-١٨)

$x_i$	$y_i$	$\hat{y}_i$	$1 - \hat{y}_i$	$\hat{w}_i = \frac{1}{\hat{y}_i(1 - \hat{y}_i)}$
2900	0	0.1962	0.8038	6.3401
7700	1	0.9305	0.0695	15.4707
3300	0	0.2574	0.7426	5.2313
4500	1	0.4410	0.5590	4.0565
5900	0	0.6552	0.3448	4.4263
7700	0	0.9305	0.0695	15.4707
3800	1	0.3339	0.6661	4.4961
3600	0	0.3033	0.6967	4.7322
5100	1	0.5328	0.4672	4.0173
7500	1	0.8999	0.1001	11.1053
6200	1	0.7011	0.2989	4.7716
3400	0	0.2727	0.7273	5.0417
6500	1	0.7470	0.2530	5.2907
2900	0	0.1962	0.8038	6.3401
4000	1	0.3645	0.6355	4.3170
8000	1	0.9764	0.0236	43.4519
7250	1	0.8617	0.1383	8.3909
3100	0	0.2268	0.7732	5.7020
3300	0	0.2574	0.7426	5.2313
7600	1	0.9152	0.0848	12.8904

باستخدام طريقة المربعات الصغرى المرجحة والتي تناولناها في الفصل الثاني يمكن إيجاد تقديرات لمعامل نموذج الانحدار. التقديرات لكل من  $\beta_1, \beta_0$  معطاه في جدول (٧-١٩) .

قيمة  $t$  اللازمة لاختبار فرض العدم  $H_1: \beta_1 = 0$  معطاه في جدول (٧-١٩) . يتضح من عمود  $p$ -value أننا نرفض فرض العدم أن  $\beta_1 = 0$  .

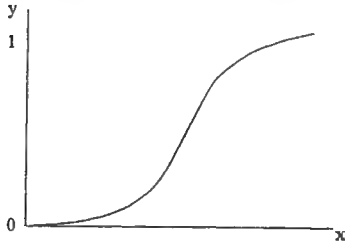
جداول (٧-١٩)

المعالم	التقدير	s.e(B <sub>i</sub> )	t	p-value المعنوية
$\beta_0$	-0.264	0.294	-0.899	0.381
$\beta_1$	0.0001	0.000	3.322	0.004

(٧-٦-٢) النموذج الغير خطي

في كثير من المشاكل عندما يعبر عن متغير الاستجابة بمتغير صوري فسي نموذج الانحدار ، تكون العلاقة بين  $Y$  ،  $x$  غير خطية وغالبا ما يأخذ نموذج الانحدار شكل  $S$  كما هو موضح في شكل (٧-٨). يوجد عدة طرق لايجاد معادلة الانحدار المقدره لنموذج الانحدار. في هذا الجزء سوف نقدم واحدة من الطرق وذلك باستخدام الدالة اللوجستية logistic function التالية:

$$\mu_{y|x} = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad (٧-٤)$$



شكل (٧-٨)

عادة تطبيق الدالة (٧-٤) في مجال الهندسة والاعمال.

ان ميزه هذه الداله هي سهوله جعلها خطيه. ويمكن جعل الداله اللوجستيه خطية بإستخدام التحويلة التالية :

$$p^* = \ln \left[ \frac{\mu y | x}{1 - \mu y | x} \right] = \ln \left[ \frac{p}{1 - p} \right]$$

وعلى ذلك :

$$p^* = \beta_0 + \beta_1 x. \quad (٧-٥)$$

وعادة لإيجاد معادلة الانحدار المقدره لدالة الاستجابة اللوجستيه (٧-٥) سوف نفترض ان هناك تكرارات من  $y$  لكل مستوى من مستويات  $x$ . سوف نرمز لمستويات  $x \rightarrow x_1, x_2, \dots, x_m$ . ويفرض أن هناك  $n_i$  من قيم المشاهدات  $y$  قد تكرر لكل مستوى من مستويات  $x$  (حيث  $y$  تأخذ القيمة 1 أو 0  $n_i$  من المرات لكل مستوى من مستويات  $x$ ). فإذا فرضنا أن  $r_i$  من عدد مرات ظهور الـ 1 لكل مستوى من مستويات  $x$  فإن نسبة ظهور الرقم 1 في كل مستوى من مستويات  $x$  هو:

$$\bar{p}_i = \frac{r_i}{n_i}, \quad i = 1, 2, \dots, n.$$

يمكن توفيق دالة الاستجابة المحوله (٧-٥) بطريقة المربعات الصغرى وذلك بإستخدام القيم التالية:

$$\bar{p}_i^* = \ln \left( \frac{\bar{p}_i}{1 - \bar{p}_i} \right).$$

كقيم لمشتغير الاستجابة .

في بعض الأحيان فإن حدود الخطأ في النموذج الخطي (النموذج المحول) ليس لها تباين متساوي . في الحقيقة عندما يكون عدد المشاهدات عند كل مستوى من  $x$  كبير فإن تباين  $\bar{p}_i^*$  سوف يكون تقريباً يساوي:

$$\text{Var}(\bar{p}_i^*) = \frac{1}{n_i \bar{p}_i (1 - \bar{p}_i)}, \quad i = 1, 2, \dots, m.$$

ولأن تباين حدود الخطأ غير ثابت فإننا يجب استخدام طريقة المربعات الصغرى المرجحة لتقدير  $\beta_0, \beta_1$  - الاختيار المناسب للوزان هو:

$$w_i = n_i \bar{p}_i (1 - \bar{p}_i) , \quad i = 1, 2, \dots, n.$$

نموذج الانحدار المقدر سوف يكون:

$$\hat{p}^* = b_0 + b_1 x .$$

حيث  $b_0, b_1$  تم الحصول عليها باستخدام طريقة المربعات الصغرى المرجحة.

معادلة الانحدار المقترحة للنموذج (٧-٤) ستكون :

$$\hat{p} = \frac{\exp(b_0 + b_1 x)}{1 + \exp(b_0 + b_1 x)} .$$

هناك مناقشة جيدة للنموذج اللوجستي مقدمه من قبل Myers(1990) .

مثال (٧-٥)

تم إعطاء تراكيز مختلفة من دواء معين لعينات مختلفة من المرضى المصابين بمرض معين والنتائج معطاه في جدول (٧-٢٠).

جدول (٧-٢٠)

تراكيز الدواء $x$ (i.m)	حجم العينة $n_i$	الاستجابة (1=يشفى 0=لم يشف) $y$	عدد المرضى $r_i$ الذين تم شفائهم
1.0	50	1,0,0,0,1,1,1,1,0,...	5
2.0	60	1,1,1,0,1,0,0,0,1,...	8
3.0	100	0,0,1,0,1,1,1,0,...	15
4.0	120	1,1,0,1,0,1,0,1,1,...	20
5.0	80	0,0,0,1,1,1,0,0,1,...	20
6.0	70	1,0,1,1,0,0,1,1,...	25
7.0	80	1,0,0,1,1,1,1,0,...	30
8.0	80	1,1,1,0,0,1,0,0,...	35
9.0	80	1,0,0,0,1,1,1,1,...	42
10.0	80	0,1,1,1,0,1,0,1,1,...	45

وخطوات الحل سوف تكون كالتالي:

- نحسب  $\bar{p}_i$  و  $(1 - \bar{p}_i)$  لكل مستوى من مستويات الدواء.  
فمثلا نسبة المرضى الذين تم شفائهم عند اعطائهم الدواء بتركيز 1.0 هي:

$$\bar{p}_1 = \frac{r_1}{n_1} = \frac{5}{50} = 0.10$$

أذن :

$$(1 - \bar{p}_1) = 1 - 0.10 = 0.90$$

وهكذا .

- نحسب قيمة  $\frac{\bar{p}_i}{1 - \bar{p}_i}$  . فمثلا للدواء ذو التركيز 1.0 تكون النسبة:

$$\frac{\bar{p}_1}{1 - \bar{p}_1} = \frac{0.10}{.90} = 0.111$$

- نحسب قيمة التحويل اللوغاريتمي :

$$\bar{p}_1^* = \ln\left(\frac{\bar{p}_1}{1 - \bar{p}_1}\right) .$$

فمثلا للدواء ذو التركيز 0.10 تكون قيمة التحويل:

$$\bar{p}_1^* = \ln(0.1111) = -2.1972 .$$

- نحسب قيمة الوزن  $w_i$  حيث ان  $w_1$  مثلا هي:

$$\begin{aligned} w_1 &= n_1 \bar{p}_1 (1 - \bar{p}_1) \\ &= 50(0.1000)(0.90) = 4.5 . \end{aligned}$$

والنتائج معطاه في جدول (٧-٢١) .

جدول (٧-٢١)

تركيز الدواء x	n <sub>i</sub>	r <sub>i</sub>	$\bar{p}_i = \frac{r_i}{n_i}$	(1 - $\bar{p}_i$ )	$\left( \frac{\bar{p}_i}{1 - \bar{p}_i} \right)$	$\bar{p}^* = \ln \left( \frac{\bar{p}_i}{1 - \bar{p}_i} \right)$	w <sub>i</sub> = n <sub>i</sub> $\bar{p}_i$ (1 - $\bar{p}_i$ )
1.0	50	5	0.1000	0.9000	0.1111	-2.1972	4.5000
2.0	60	8	0.1333	0.8667	0.1538	-1.8718	6.9333
3.0	100	15	0.1500	0.8500	0.1765	-1.7346	12.7500
4.0	120	20	0.1667	0.8333	0.2000	-1.6094	16.6667
5.0	80	20	0.2500	0.7500	0.3333	-1.0986	15.0000
6.0	70	25	0.3571	0.6429	0.5556	-0.5878	16.0714
7.0	80	30	0.3750	0.6250	0.6000	-0.5108	18.7500
8.0	80	35	0.4375	0.5625	0.7778	-0.2513	19.6875
9.0	80	42	0.5250	0.4750	1.1053	0.1001	19.9500
10.0	80	45	0.5625	0.4375	1.2857	0.2513	19.6875

وبتطبيق طريقة المربعات الصغرى المرجحة نحصل على معادلة الانحدار المقدرة :

$$\hat{p}^* = -2.556 + 0.290x .$$

ولاعادة التحويل الى الارقام الاصلية نتبع الآتي:

أولاً: نحسب  $\hat{p}_1^*$  ثم بعد ذلك اعادة التحويل إلى الارقام الاصلية فمثلاً لتركيز الدواء 1.0 فإننا نجد القيمة المتوقعة  $\hat{p}_1^*$  وهي:

$$\begin{aligned} \hat{p}_1^* &= -2.556 + 0.290(1.0) \\ &= -2.266. \end{aligned}$$

والآن نحول  $\hat{p}_1^*$  الى قيمتها الاصلية كالآتي:

$$\begin{aligned} \hat{p}_1 &= \frac{\exp(b_0 + b_1 x)}{1 + \exp(b_0 + b_1 x)} \\ &= \frac{e^{\hat{p}_1^*}}{1 + e^{\hat{p}_1^*}} = \frac{\bar{e}^{-2.266}}{1 + \bar{e}^{-2.266}} \\ &= 0.09398. \end{aligned}$$

والتعبير للانحراف المعياري لـ  $B_i$  و  $i = 0, 1$  (s.e( $B_i$ )) معطاة في جدول (٧-٢٢) .

يعطي جدول (٧-٢٢) قيمة  $t$  اللازمة لاختبار فرض العدم  $H_1: \beta_1 = 0$ .  
يتضح من عمود  $p$ -value أننا نرفض فرض العدم أن  $\beta_1 = 0$ .

جدول (٧-٢٢)

المعالم	التقدير	الخطأ المعياري $se(B_i)$	$t$	$p$ -value المعنوية
$\beta_0$	-2.556	0.108	-23.639	0.000
$\beta_1$	0.290	0.016	18.473	0.000



الفصل الثامن  
الارتباط الذاتي  
**Autocorrelation**

- (١-٨) مقدمة
- (٢-٨) أسباب الارتباط الذاتي
- (٣-٨) اختبار درين \_ واتسون
- (٤-٨) معالجة الارتباط الذاتي
- (١-٤-٨) الطريقة الاولى
- (٢-٤-٨) الطريقة الثانية
- (٣-٤-٨) الطريقة الثالثة
- (٤-٤-٨) الطريقة الرابعة

### (٨-١) مقدمته

علماً فيما سبق أنه لتقدير معالم نموذج الإحداد الخطي فيجب تحقق الفروض التالية لحدود الخطأ :

$$E(\varepsilon_i) = 0 \quad \text{Var}(\varepsilon_i) = \sigma^2 \quad E(\varepsilon_i \varepsilon_j) = 0, \quad i \neq j$$

لفرض إختبارات الفروض والحصول على فترات ثقة عادة يضاف فرض الإعتدال أي أن :  $\varepsilon_i \sim \text{NID}(0, \sigma^2)$  . بعض تطبيقات الإحداد تستعمل على متغيرات مستقلة ومتغير إستجابة يكون له طبيعة للتتابع مع الزمن و البيانات في هذه الحالة تسمى السلاسل الزمنية . معظم المسائل الاقتصادية تكون على شكل سلاسل زمنية مما يؤدي إلى أن الخطأ في فترة زمنية  $\varepsilon_i$  يكون مرتبطاً مع الخطأ  $\varepsilon_j$  في فترة زمنية أخرى ( $i \neq j$ ) وهذا يخالف إحدى فروض نموذج الإحداد الخطي وهو عدم إرتباط قيمة  $\varepsilon$  في فترة زمنية ما عن قيمتها في فترة زمنية سابقة ، أي أن الإرتباط بين  $\varepsilon_i, \varepsilon_j$  لا يساوي الصفر ( $E(\varepsilon_i \varepsilon_j) \neq 0$ ) . ومعنى ذلك وجود الإرتباط الذاتي أو الإرتباط التسلسلي . والإرتباط الذاتي حالة خاصة من الإرتباط إذ يقسم لنا درجة الإرتباط بين القيم المتتالية لنفس المتغير خلال فترة زمنية محددة وليس بين متغيرين أو أكثر . وستقتصر دراستنا هنا فقط على الحالة البسيطة وهي حالة العلاقة الخطية بين إبي قيمتين متتاليتين من قيم  $\varepsilon$  حيث :

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i$$

حيث  $u_i$  متغير عشوائي (يسمى حد الإضطراب) يتبع التوزيع الطبيعي بمتوسط يساوي صفر وتباين  $\sigma_u^2$  و  $i \neq j$  ,  $E(u_i u_j) = 0$  و معامل الارتباط الذاتي البسيط حيث  $| \rho | < 1$  . وتعرف العلاقة السابقة بأنها إحداد ذاتي

(autoregressive) من الدرجة الأولى ، وسنبدأ التحليل بصيغة العلاقة البسيطة بين المتغيرات العشوائية ، وبمعنى آخر سنبدأ بمعامل الإرتباط الذاتي البسيط  $\rho$  كعلاقة خاصة لمعامل الإرتباط البسيط  $r$  . والمعاملان يتشابهان من حيث أن كليهما لا يناسب العلاقات الغير خطية ، وإن  $\rho$  يكون مناسباً فقط إذا كانت قيمة  $\varepsilon_i$  في نقطة زمنية تتبع قيمتها في النقطة الزمنية التي تسبقها فقط . أما إذا كانت قيم  $\varepsilon_i$  تتبع قيمة النقطتين السابقتين فإن صيغة العلاقة تكون من نوع الإحداد الذاتي من الدرجة الثانية . والطريقة المستخدمة في بحوث الاقتصاد القياسي للتطبيقي للتعرف على وجود الإرتباط الذاتي هو رسم البواقي المعاوية مقابل الزمن فإذا أخذت هذه البواقي شكلاً منتظماً كالأمثان أو الدورات أكد ذلك وجود الإرتباط

الذاتي بين الأخطاء . وتتحدد إشارة معامل الارتباط الذاتي حسب تغير إشارة قيم البواقي ، فإذا تغيرت إشارة القيم المتتالية باستمرار فيأخذ المنحنى التاريخي شكل الأسنان كان الارتباط سالبا ، والعكس إذا حدث التغير بأن يتلو عددا من القيم الموجبة عددا آخر من القيم السالبة كان الارتباط موجبا .

إذا كانت حدود الخطأ في نموذج الإنحدار مرتبطة ارتباط ذاتي موجب ، فإن استخدام طريقة المربعات الصغرى يترتب عليه عدد من العواقب المهمة وهي :

١- لا تزال معاملات الإنحدار المقدرة غير متحيزة إلا أنها لا تمتلك خاصية أقل تباين .

٢- متوسط مربعات الخطأ يمكن أن يشكل تقديرا بالنقصان لتباين حدود الخطأ .

٣- تعطى التقديرات للأخطاء المعيارية لمقدرات معاملات الإنحدار ،

$s.e(B_i), i = 0, 1, 2, \dots, k$  ، والمحسوبة بطريقة المربعات الصغرى تقديرا بالنقصان للانحراف المعياري الحقيقي للمقدر  $B_i$  .

٤- لم تعد فترات الثقة والاختبارات التي تستخدم توزيعات  $t$  أو  $F$  قابلة للتطبيق .

### (٢-٨) أسباب الارتباط الذاتي

تبدأ مشكلة الارتباط الذاتي في بيانات السلاسل الزمنية بطبيعة البيانات نفسها وطرق تجميعها. فقد يترتب على وجود أخطاء القياس في تجميع هذا النوع من البيانات أخطاء تراكمية في السنوات أو النقاط الزمنية التالية. يلي ذلك الصيغة الدالية المستخدمة لتقدير النموذج المستخدمة بالإضافة إلى عدم إجراء التحويلات

المناسبة للمتغيرات لجعل نموذج الإنحدار خطي في المعالم. وكذلك قد يؤدي إغفال إدخال متغيرات في الدالة إلى وجود ارتباط ذاتي. ومن العوامل التي تؤدي إلى وجود ارتباط ذاتي البيانات المستكملة نتيجة لاستخدام بيانات تعدادات على فترات زمنية متباعدة وهذا ما يتم في بيانات تعداد السكان حيث لا يجرى الإجراء الإحصاء الفعلي كل عام وإنما كل خمس سنوات أو عشر سنوات ثم تقدر قيمة السنوات بين هذه الفترات وكذلك هو الحال حين يتم استكمال نوع ما من البيانات للتعرض عن قيم مفقودة. توجد طرق كثيرة لاكتشاف عدم استقلال الأخطاء وسوف تقتصر دراستنا على اختبار درين \_ واتسون.

### (٣-٨) اختبار درين \_ واتسون

إن النموذج الخطي (١-١) في وجود ارتباط ذاتي من الرتبة الأولى

هو:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (١-٨)$$

حيث:

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i \quad (٢-٨)$$

حيث  $\rho$  معامل الارتباط الذاتي بحيث  $|\rho| < 1$  و  $u_i$  متغير عشوائي يتبع التوزيع الطبيعي بمتوسط يساوي صفر وتباين ثابت  $\sigma_u^2$  و  $E(u_i u_j) = 0, i \neq j$ .

يستخدم اختبار درين \_ واتسون لاختبار ثلاثة فروض وهي :

١- وجود ارتباط ذاتي موجب :

$$H_0 : \rho = 0$$

ضد الفرض البديل :

$$H_0 : \rho > 0$$

٢- وجود ارتباط ذاتي سالب :

$$H_0 : \rho = 0$$

ضد الفرض البديل :

$$H_0 : \rho < 0$$

٣- وجود ارتباط ذاتي سالب أو موجب (اختبار ذو جانبيين) :

$$H_0 : \rho = 0$$

ضد الفرض البديل :

$$H_1 : \rho \neq 0$$

وينحصر الاختبار بالخطوات التالية:

أ- تقدير معالم الانحدار باستخدام أسلوب المربعات الصغرى للحصول على معاملات الانحدار.

ب- طرح قيم المتغير التابع من القيم المشاهدة للحصول على البواقي:

$$e_i = y_i - \hat{y}_i.$$

ج- حساب قيمة إحصائية مقدرة ترمز لها بالرمز DW على النحو التالي:

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}.$$

مع ملاحظة أن:

$$0 \leq DW \leq 4.$$

د- استخدام جداول درين \_ واتسون في الملحق (٨) لإجراء الاختبار ومن الملاحظ أن جداول درين \_ واتسون تأخذ في الاعتبار كل من عدد المشاهدات  $n$  وعدد المتغيرات المستقلة ( $k$ ) ومستوى المعنوية  $\alpha$  في حالة اختبار من جانب واحد و  $2\alpha$  في حالة اختبار ذو جانبيين . ومما هو جدير بالذكر أن للفرض الأكثر شيوعا هو الفرض البديل :  $H_1: \rho > 0$  ويحتوي الجدول على قيمتين إحداهما  $d_L$  وهي للقيمة الصغرى و  $d_U$  العليا ثم تتم المقارنة على النحو التالي الموضح في جدول (٨-١).

جدول (٨-١)

الحالة	قيمة DW المقدره	القرار
1	$4 - d_L < DW < 4$	ارتباط ذاتي سالب
2	$4 - d_U < DW < 4 - d_L$	قرار غير محدد
3	$2 < DW < 4 - d_U$	لا يوجد ارتباط ذاتي
4	$d_U < DW < 2$	لا يوجد ارتباط ذاتي
5	$d_L < DW < d_U$	قرار غير محدد
6	$0 < DW < d_L$	ارتباط ذاتي موجب

مما تقدم نجد أن هناك ثلاث نتائج للاختبار:

١. لا يوجد ارتباط ذاتي في الحالتين ٣، ٤ .

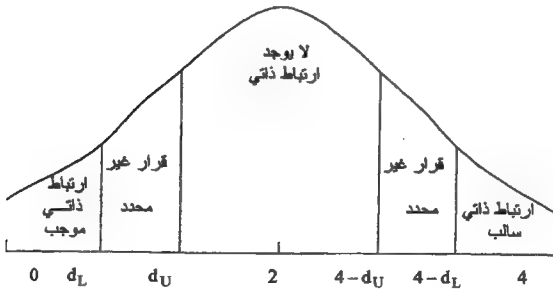
٢. قرار غير محدد أي لا يمكن الجزم بوجود أو عدم وجود ارتباط ذاتي وذلك يستلزم إضافة بيانات إلى السلسلة الزمنية إن أمكن كما في الحالتين ٢، ٥.

٣. وجود ارتباط ذاتي سالب كما في الحالة الاولى أو وجود ارتباط ذاتي موجب كما في الحالة السادسة.

يوضح شكل (١-٨) مناطق اتخاذ القرار مبينة عليه قيم  $DW$  و  $d_L$  ,  $d_U$  .  
ويجب حساب معامل الارتباط البسيط بين الأخطاء وبعضها  $\hat{\rho}$  وذلك فسي  
الحالة التي يوجد فيها ارتباط ذاتي نتيجة الاختبار وتوجد معادلة تقريبية لحساب  $\hat{\rho}$   
من  $DW$  على النحو التالي:

$$\hat{\rho} = 1 - DW/2.$$

وقد سبق أن ذكرنا أن قيمة  $DW$  تتراوح بين الصفر وأربعة. إذا كانت  $DW=0$   
فهذا يعني أن  $\hat{\rho} \rightarrow 1$  وإذا كانت  $DW=4$  فهذا يعني أن  $\hat{\rho} \rightarrow -1$  .  
أي أنه إذا اقتربت قيمة  $DW$  من الصفر نجد أن هناك ارتباط ذاتيا موجبا وكلما  
اقتربت قيمة  $DW$  من 4 سنجد أن هناك ارتباطا ذاتيا عكسيا.



شكل (١-٨)

مثال (١-٨)

تبين البيانات في جدول (٢-٨) قيم لمتغيرين  $Y$  و  $X$  ناتجة من عينة عشوائية.  
المطلوب إجراء اختبار للارتباط الذاتي مستخدما مستوى معنوية  $\alpha = 0.05$   
وأذكر الفرضيات البديلة وقاعدة القرار والنتيجة.

جدول (٢-٨)

x	18	14	10	15	7	12	13	8	9	17	15	12
y	20	11	14	16	10	10	17	11	12	20	18	12

الحل

أولاً: لإختبار فرض العدم  $H_0: \rho = 0$  ضد الفرض البديل  $H_1: \rho > 0$  يتم إيجاد تقديرات معالم نموذج الانحدار البسيط بطريقة المربعات الصغرى حيث:

$$\bar{y} = \frac{\sum y_i}{n} = \frac{171}{12} = 14.25$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{150}{12} = 12.5$$

$$b_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{2255 - \frac{(171)(150)}{12}}{2010 - \frac{(171)^2}{12}} = \frac{117.5}{135} = 0.87037 ,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 3.37037$$

وبالتالي فإن معادلة الانحدار المقدرة سوف تكون :

$$\hat{y} = 3.37037 + 0.87037x.$$

يعطي جدول (٢-٨) القيم اللازمة لحساب قيمة لاحصاء درين \_ واتسون . ومن القيم الواردة بالجدول (٢-٨) يتم حساب DW على النحو التالي :

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} = \frac{150.794}{55.9815} = 2.69363$$

قيمتي  $d_U, d_L$  من جدول درين- واتسون في ملحق (٨) عند مستوي معنوية  $\alpha = 0.05$  و  $k = 1, n = 15$  ( لعدم وجود قيمة عند  $n=12$  ) هي:  $d_U = 1.36, d_L = 1.08$  وبالتالي فإن مناطق القبول يمكن ملاحظتها من شكل

(٨-١) . نلاحظ أن قيمة  $DW$  تقع بين  $4 - d_L$  و  $4 - d_U$  أي بين  $4 - 1.08$  و  $4 - 1.36$ ، أي بين  $2.92$  و  $2.64$  في منطقة قرار غير محدد أي أنه لا يمكن الحكم على قيمة  $DW$ . في هذه الحالة فإننا نحتاج إلى مزيد من المشاهدات. وفي حالة بيانات سلاسل زمنية قد يكون من المستحيل، بالطبع، الحصول على مزيد من البيانات أو من الممكن أن تتوافر المشاهدات المطلوبة في المستقبل مما يؤدي إلى تأخير كبير عند الانتظار للحصول عليها.

جدول (٨-٣)

$y_i$	$x_i$	$\hat{y}_i$	$e_i$	$e_{i-1}$	$(e_i - e_{i-1})^2$	$e_i^2$
20	18	19.0376	0.9624	-	-	0.9273
11	14	15.556	-4.556	0.9624	30.4527	20.757
14	10	12.0744	1.9256	-4.556	42.0111	3.7079
16	15	16.426	-0.4264	1.9256	5.5319	0.1818
10	7	9.4632	0.5360	-0.4264	0.9278	0.2882
10	12	13.8152	-3.8152	0.5368	18.7379	14.5558
17	13	14.6856	2.3144	-3.8152	37.5770	5.3564
11	8	10.336	0.664	2.3144	2.7238	0.4409
12	9	11.204	0.796	0.664	0.0174	0.6336
20	17	18.1672	1.8328	0.796	1.0750	3.3592
18	15	16.4264	1.5736	1.832	0.0672	2.4762
12	12	13.8152	-1.8252	1.5736	11.4840	3.2950



#### (٤-٨) معالجة الارتباط الذاتي

سنناقش فيما يلي أهم الطرق للتخلص من وجود الارتباط الذاتي بين الأخطاء ، وكما سبقت الإشارة إلى أسباب وجود الارتباط الذاتي فإنه يرجع إلى عدة عوامل منها الدالة المستخدمة وإغفال بعض المتغيرات وسنحاول التأكد من ذلك من خلال استخدام الطرق المختلفة للتخلص من الارتباط الذاتي.

ويجب أن ننوه هنا أن بعض الطرق المستخدمة سوف تقدم في حالة الانحدار الخطي البسيط وذلك للتسهيل ويمكن تعميم تلك الطرق في حالة الانحدار الخطي المتعدد .

#### (١-٤-٨) الطريقة الأولى

غالباً يلجأ الاقتصاديون لتسهيل الأمور بإدخال المتغير التابع كمتغير مستقل بمعنى آخر إذا كانت السلسلة تبدأ من 1992 للمتغير التابع . وكذلك المتغيرات المستقلة فإنه يمكن إدخال بيانات المتغير التابع لعام 1991 مقابل المتغير التابع لعام 1992 وتسمى هذه الطريقة *lagged variable* أي متغير إبطاء. ولكن لا تتوفر في معظم الأحيان البيانات عن عام سابق للسلسلة الزمنية المستخدمة. في هذه الحالة يمكن التضحية بمشاهدة واحدة نظير التخلص من أثر الارتباط الذاتي ويوصف النموذج على النحو التالي:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 Y_{i-1} + \varepsilon_i.$$

#### مثال (٢-٨)

يعطى جدول (٤-٨) البيانات عن دولة الكويت للأعوام خلال الفترة (1986-1962) متضمنة :

$x$  = الدخل المتاح أو الدخل الذي يمكن للتصرف به.

$y$  = الاستهلاك الخاص.

**والمطلوب :** تقدير معادلة الانحدار البسيط مع إجراء اختبار وجود ارتباط ذاتي بين الأخطاء ثم استخدام الطريقة الأولى للتخلص من وجود الارتباط الذاتي بين الأخطاء

جدول (٤-٨)

x	y	السنة
460.0	188.0	1962
486.0	192.0	1963
561.0	200.0	1964
553.0	191.0	1965
682.0	232.0	1966
1010.0	280.0	1967
793.0	297.0	1968
840.0	306.0	1969
851.0	396.0	1970
1117.0	420.0	1971
1102.0	227.0	1972
1262.0	439.0	1973
3532.0	564.0	1974
3711.0	759.0	1975
4281.0	1030.0	1976
4563.0	1368.0	1977
4977.0	1474.0	1978
7597.0	1671.0	1979
8757.0	2196.0	1980
8875.0	2445.0	1981
7612.0	3287.0	1982
7789.0	3179.0	1983
7893.0	2780.0	1984
7322.0	2774.0	1985
7164.0	2575.0	1986

الحل
------

تم الحصول على قيمة DW باستخدام برنامج SPSS حيث  $DW = 0.86467$  ولإجراء اختبار درين \_ واتسون يجب الحصول على قيم  $d_L, d_U$  المناظرة لمتغير واحد ( $k=1$ ) وعدد المشاهدات  $n = 25$  ومستوى معنوية  $\alpha = 0.05$  حيث  $d_U = 1.45, d_L = 1.20$  ومنه نجد أن:

$$0 < DW < d_L$$

أي أن :

$$0 < 0.86467 < 1.20$$

وبذلك نرفض فرض العدم :

$$H_0 : \rho = 0$$

ونقبل الفرض البديل :

$$H_1 : \rho > 0$$

ويمكن حساب  $\hat{\rho}$  من العلاقة التقريبية التالية :

$$\begin{aligned}\hat{\rho} &= 1 - \frac{DW}{2} = 1.0 - 0.86467/2 \\ &= 0.56767\end{aligned}$$

ونلك يعني أن معامل الارتباط بين الأخطاء يساوي 0.56767 لذا لا يمكن الاعتماد على النتائج التي نحصل عليها من معادلة الانحدار المقترنة:

$$\hat{y} = -30.43725 + 0.32233x$$

ولا يمكن استخدامها للتنبؤ قبل تخلصها من وجود الارتباط الذاتي بين الأخطاء. الآن للتخلص من وجود الارتباط الذاتي بين الأخطاء سوف نستخدم الطريقة الأولى.

من البيانات في جدول (8-5) سوف نستخدم متغير إبطاء للتخلص من وجود الارتباط الذاتي مع التحقق من ذلك بإجراء اختبار درين \_ واتسون على الأخطاء المقترنة طبقا للنموذج الجديد.

جدول (٨-٥)

$x_i$	$y_i$	$(z) y_{i-1}$	السنة
460.0	188.0	-	1962
486.0	192.0	188.0	1963
561.0	200.0	192.0	1964
553.0	191.0	200.0	1965
682.0	232.0	191.0	1966
1010.0	280.0	232.0	1967
793.0	297.0	280.0	1968
840.0	306.0	297.0	1969
851.0	396.0	306.0	1970
1117.0	420.0	396.0	1971
1102.0	227.0	420.0	1972
1262.0	439.0	227.0	1973
3532.0	564.0	439.0	1974
3711.0	759.0	564.0	1975
4281.0	1030.0	759.0	1976
4563.0	1368.0	1030.0	1977
4977.0	1474.0	1368.0	1978
7597.0	1671.0	1474.0	1979
8757.0	2196.0	1671.0	1980
8875.0	2445.0	2196.0	1981
7612.0	3287.0	2445.0	1982
7789.0	3179.0	3287.0	1983
7893.0	2780.0	3179.0	1984
7322.0	2774.0	2780.0	1985
7164.0	2575.0	2774.0	1986

معادلة الانحدار المقدره سوف تكون

$$\hat{y} = 4.2508 + 0.11354x + 0.69095z$$

حيث  $z$  هو متغير إخطاء. قيمة  $DW$  هي 2.23556 وقيم  $d_U, d_L$  عندما  $\alpha = 0.05, n = 24$  ومتغيرين مستقلين هما (تقريباً):  $d_U = 1.45, d_L = 1.2$  و

$$2 < DW < (4 - d_U)$$

$$2 < 2.23556 < 2.55$$

وبذلك نضمن عدم وجود ارتباط ذاتي بين الأخطاء ونقبل فرض العدم  $H_0: \rho = 0$  حيث  $\hat{\rho} = 1 - DW/2 = -0.117$  التقريبية وهو ارتباط ضئيف بين الأخطاء مما يؤكد القرار.

#### (٨-٤-٢) الطريقة الثانية

سبق أن ذكرنا أن أهمل أحد المتغيرات قد يؤدي إلى وجود الارتباط الذاتي. في هذه الطريقة يتم توصيف الدالة وإدخال متغيرات ثم إهمالها. في المثال (٨-٢) ولما كانت الفترة 1982-1986 تتصف بتراجع الدخل مع زيادة الاستهلاك نتيجة لعوامل خارجية منها انخفاض اسعار النفط الخام وآثار أزمة "سوق المناخ" فيمكن إدخال متغير صوري  $w$  تكون قيمة مساوية الواحد الصحيح خلال الفترة 1986-1982 ومساوية للصفر فيما عدا ذلك والبيانات اللازمة لإيجاد معادلة الانحدار الجديدة معطاه في جدول (٨-٦).

جدول (٨-٦)

x	y	السنة	w
460.0	188.0	1962	0
486.0	192.0	1963	0
561.0	200.0	1964	0
553.0	191.0	1965	0
682.0	232.0	1966	0
1010.0	280.0	1967	0
793.0	297.0	1968	0
840.0	306.0	1969	0
851.0	396.0	1970	0
1117.0	420.0	1971	0
1102.0	227.0	1972	0
1262.0	439.0	1973	0
3532.0	564.0	1974	0
3711.0	759.0	1975	0
4281.0	1030.0	1976	0
4563.0	1368.0	1977	0
4977.0	1474.0	1978	0
7597.0	1671.0	1979	0
8757.0	2196.0	1980	0
8875.0	2445.0	1981	0
7612.0	3287.0	1982	1
7789.0	3179.0	1983	1
7893.0	2780.0	1984	1
7322.0	2774.0	1985	1
7164.0	2575.0	1986	1

معادلة الاتحاد المقتره سوف تكون على الشكل:

$$y = 61.131 + 0.24375x + 1016.102w$$

قيمة DW هي 1.55, 1.36727,  $d_L = 1.21$  , ونلك عند  
 $\alpha = 0.05$  ,  $k = 2$  ,  $n = 25$  وبناء عليه نجد أن:

$$d_L < DW < d_U$$

$$1.21 < DW < 1.55$$

$$1.21 < 1.36727 < 1.55$$

وبالتالي لا يوجد ارتباط ذاتي بين الأخطاء ونقبل فرض العدم  $H_0: \rho = 0$ .

#### (٣-٤-٨) الطريقة الثالثة

اتضح من الطريقتين المابقتين ان تجاهل أو إغفال بعض المتغيرات في  
 توصيف النموذج ادى بدوره إلى وجود ارتباط ذاتي للباقي في النموذج الأصلي  
 وبمعالجة النموذج للتخلص من الارتباط الذاتي وجد أن كلتا الطريقتين أدت إلى  
 تحسين DW لتقع في منطقة قبول فرض العدم  $H_0: \rho = 0$ . والطريقة التالية  
 تستخدم ما يطلق عليه الطريقة العامة للمربعات الصغرى. وتعتمد هذه الطريقة على  
 تحويل البيانات الأصلية إلى الصورة التي تمكنا من الحصول على نموذج يكون  
 المتغير العشوائي فيه خاضع لفروض طريقة المربعات الصغرى العادية وبالتالي  
 يمكن استخدام هذه الطريقة في تقدير المعالم.

بفرض النموذج:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (٣-٨)$$

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i \quad \text{و} \quad |\rho| \leq 1$$

حيث:

$$u_i \sim N(0, \sigma_u^2), E(u_i u_j) = 0, i \neq j$$

ثم بكتابة النموذج السابق في الملاحظة i-1 نحصل على:

$$Y_{i-1} = \beta_0 + \beta_1 x_{i-1} + \varepsilon_{i-1}, \quad (٤-٨)$$

وبضرب طرفي (٤-٨) في  $\rho$  والطرح من (٣-٨) نحصل على النموذج التالي:

$$Y_i^* = \beta_0(1-\rho) + \beta_1 x_i^* + u_i, \quad (٥-٨)$$

حيث:

$$Y_i^* = Y_i - \rho Y_{i-1}, \quad (٦-٨)$$

$$x_i^* = x_i - \rho x_{i-1}, \quad (٧-٨)$$

$$u_i = \varepsilon_i - \rho \varepsilon_{i-1}$$

النموذج المحول (٥-٨) يصبح :

$$Y_i^* = \beta_0^* + \beta_1^* x_i^* + u_i \quad (٨-٨)$$

$$\beta_0^* = \beta_0(1-\rho) \quad , \quad \beta_1^* = \beta_1 \quad \text{حيث}$$

وعلى ذلك :

$$\beta_0 = \beta_0^* / (1-\rho)$$

وبذلك أمكن تحويل النموذج الذي يحتوي على الارتباط الذاتي إلى نموذج لا يحتوي على ارتباط ذاتي بين البواقي وبذلك يمكن استخدام طريقة المربعات الصغرى العادية لاشتقاق تقديرات المعامل وهي نفس معالم النموذج الأصلي ماعدا المعطية  $\beta_0^* = \beta_0(1-\rho)$ . ويجب ملاحظة أن عدد المشاهدات المحولة الداخلة في التقدير هي  $n-1$  وأن المتغير العشوائي لا غير مرتبط ذاتيا. ونلاحظ أن هذه الطريقة تعتمد على معرفة قيمة معامل الارتباط الذاتي  $\rho$  ونادر ما تكون قيمة معامل الارتباط الذاتي معلومة ، وبالتالي نحتاج لتقديرها. ولحسن الحظ يوجد عدد من الطرق المستخدمة لتقدير قيمة معامل الارتباط الذاتي  $\rho$  وسوف نتناولها لاحقا .

وعلى ذلك فبعد اختبار وجود الارتباط الذاتي والحصول على تقدير لـ  $\rho$  يتم تطبيق طريقة المربعات الصغرى العادية على مجموع البيانات المحولة  $y_i^*, x_i^*$  حيث تطرح من المشاهدات الأصلية في كل نقطة زمنية حاصل ضرب  $\hat{\rho}$  في قيمة المتغيرات في الفترات السابقة كالتالي:

$$y_i^* = y_i - \rho y_{i-1},$$

$$x_i^* = x_i - \rho x_{i-1}$$

معادلة الانحدار المقدرة سوف تكون على الشكل:

$$\hat{y}^* = b_0^* + b_1^* x^*$$



حيث  $b_0 = b_0^* / (1 - \hat{\rho})$  و  $b_1 = b_1^*$

### طرق تقدير $\rho$

#### ١- طريقة درين والتسون لتقدير $\rho$

وهذه الطريقة يمكن تطبيقها لأي درجة من الانحدار وتتم هذه الطريقة كالآتي:

نبدأ بالنموذج المحول (٨-٨) حيث يكتب بتفصيل كالآتي:

$$Y_i - \rho Y_{i-1} = \beta_0(1 - \rho) + \beta_1(x_i - \rho x_{i-1}) + u_i \quad (٩-٨)$$

وبإعادة تنظيم (٩-٨) نحصل على:

$$Y_i = \beta_0(1 - \rho) + \beta_1(x_i - \rho x_{i-1}) + \rho Y_{i-1} + u_i, i = 1, 2, 3, \dots, n \quad (١٠-٨)$$

وباستخدام طريقة المربعات الصغرى لتقدير معالم هذا النموذج نحصل على تقدير لـ  $\rho$  ، أي  $\hat{\rho}$  ، والذي يساوي معامل المتباطئ  $Y_{i-1}$ .

#### ٢- طريقة كوكران اوركوت (Cochrane - Orcutt)

وذلك بالنظر إلى المعادلة (٨-٢) والمفروضه على النموذج (٨-١) على أنها إحدار عبر نقطة الأصل أي أن :

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i$$

حيث  $\varepsilon_i$  هو المتغير التابع ،  $\varepsilon_{i-1}$  هو المتغير المستقل ،  $u_i$  حد الخطأ و  $\rho$  ميل الخط عبر نقطة الأصل. وبما أن  $\varepsilon_i, \varepsilon_{i-1}$  غير معروفين فنستخدم  $\varepsilon_i, \varepsilon_{i-1}$  التي حصلنا عليها بطريقة المربعات الصغرى العادية كمتغير مستقل ومتغير تابع على الترتيب ونحصل على تقدير لـ  $\rho$  بتقدير خط مستقيم عبر نقطة الأصل من الصيغة التالية :

$$\hat{\rho} = \frac{\sum_{i=2}^n e_{i-1}e_i}{\sum_{i=2}^n e_{i-1}^2}$$

### ٣- طريقة هيلدرث-لو

هناك طريقة هيلدرث - لو لتقدير معامل الارتباط  $\rho$  وذلك بهدف استخدامها في التحويلات (٨-٦) و (٨-٧) والتي تتخذ نفس الأسلوب الذي تتخذه طريقة بوكس - كوكس لتقدير المعلمة  $\lambda$  في تحويل القوى  $Y$  بغية تحسين صلاحية نموذج الانحدار. إذ نختار في طريقة هيلدرث - لو تلك القيمة  $\rho$  التي تجعل مجموع مربعات الخطأ للبقايا لنموذج الانحدار المحصول (٨-٨) أصغر ما يمكن:

$$SSE = \sum (y_i^* - \hat{y}_i)^2 = \sum (y_i^* - b_0^* - b_1^* x_i^*)^2.$$

وتتوافر برنامج حاسب لإيجاد قيمة  $\rho$  التي تجعل SSE أصغر ما يمكن. وبصورة بدئية يمكننا أن نبحث حسابيا بتشغيل انحدارات متكررة مع قيم مختلفة لـ  $\rho$  في كل إنحدار وذلك لاستطلاع القيمة التقريبية لـ  $\rho$  التي تجعل SSE أصغر ما يمكن. وعند معرفة الفترة التي تقع فيها قيمة  $\rho$  التي تجعل SSE أصغر ما يمكن ، يمكن للبحث ضمن هذه الفترة على قيمة أكثر دقة لـ  $\rho$ . وبمجرد الحصول على قيمة  $\rho$  التي تجعل SSE أصغر ما يمكن يمكننا تحديد معادلة الانحدار المقدرة المقابلة لتلك القيمة لـ  $\rho$ .

### ٤- طريقة الفروق الأولى

وبما أن قيمة معامل الارتباط الذاتي  $\rho$  هي قيمة كبيرة في الغالب وأن SSE كدالة في  $\rho$  تكون مستقرة تماما من أجل قيم كبيرة لـ  $\rho$  حتى 1.0 فقد اقترح بعض الإحصائيين استخدام  $\rho = 1.0$  في النموذج المحصول (٨-٨) وإذا كانت  $\rho = 1$  فإن  $\beta_0^* = \beta_0(1 - \rho) = 0$  فيصبح النموذج المحصول (٨-٨) كمايلي:

$$Y_i^* = \beta_1^* x_i^* + u_i$$

حيث :

$$Y_i^* = Y_i - Y_{i-1} \quad (١١-٨)$$

$$x_i^* = x_i - x_{i-1} \quad (١٢-٨)$$

وهكذا نجد مرة أخرى أنه يمكن تقدير معالم نموذج الانحدار مباشرة بطريقة المربعات الصغرى وتركز هذه المرة على إحدار عبر نقطة الأصل. معادلة الانحدار المقدرة سوف تكون:

$$\hat{y} = b_1^* x^*$$

ويمكن تحويلها والعودة مرة أخرى إلى المتغيرات الأصلية كما يلي :

$$\hat{y} = b_0 + b_1 x$$

$$b_0 = \bar{y} - b_1^* \bar{x} \quad , \quad \text{حيث :}$$

$$\bar{b}_1 = b_1^* .$$

•-استخدام صيغ أخرى :

•-يمكن استخدام الصيغة التالية التي سبق أن تناولناها وهي:  $\hat{\rho} = 1 - DW/2$ .

•- يمكن تقدير معامل الارتباط الذاتي  $\rho$  من الصيغة التالية:

$$\hat{\rho} = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2} ,$$

مثال (٨-٣)

يعطي جدول (٨-٧) بيانات الواردات والنتائج القومي بالمليون جنيه في بلد ما والمطلوب اختبار الارتباط الذاتي ومعالجته.

جدول (٨-٧)

المرتبة	الواردات $y_i$	النتيجة القومية $x_i$	$\hat{y}_i$	$y_i - \hat{y}_i$	$e_i - e_{i-1}$
1950	3748	21777	3615.5	132.42	-
1951	4010	22418	3795.2	214.73	82.3
1952	3711	22308	3764.4	-53.42	-268.16
1953	4004	23319	4047.8	-43.84	9.58
1954	4151	24180	4289.2	-138.2	-94.37
1955	4469	24893	4489.08	-20.08	118.12
1956	4582	25310	4605.9	-23.98	-3.90
1957	4697	25799	4743.07	-46.07	-22.08
1958	4753	25886	4767.4	-14.46	31.61
1959	5062	26868	5042.7	19.25	33.71
1960	5669	28134	5397.6	271.35	252.10
1961	5628	29091	5665.9	-37.92	-309.28
1962	5736	29450	5766.5	-30.56	7.36
1963	5946	30705	6118.3	-172.38	-141.82
1964	6501	32372	6585.7	-84.70	87.68
1965	6549	33152	6804.3	-255.36	-170.66
1966	6705	33764	6975.9	-270.92	-15.56
1967	7104	34411	7157.3	-53.30	217.62
1968	7609	35429	7442.6	166.31	219.62
1969	8100	36200	7658.8	441.18	274.86

بإستخدام طريقة المربعات الصغرى العادية فإن معادلة الإتحدار المقدرة هي :

$$\hat{y} = -2489.25 + 0.28x$$

وإذا كان:

$$\sum e_i^2 = 567861$$

$$\sum (e_i - e_{i-1})^2 = 491847$$

وبتطبيق اختبار درين\_ واتسون فإن:

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} = \frac{491847}{567861} = 0.866$$

وبالرجوع للجدول في ملحق (٨) عند مستوى معنوية  $\alpha = 0.05$  وعدد مشاهدات 20 ومتغير مستقل واحد ( $k=1$ ) فإن  $d_L = 1.2$  ,  $d_U = 1.41$  ولما كانت  $DW < d_L$  فمن الواضح وجود ارتباط ذاتي موجب ولمعالجة الارتباط الذاتي لحسب أولا معامل الارتباط الذاتي حيث:

$$\hat{\rho} = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2} = \frac{215848}{567861} = 0.380107$$

وبتطبيق الطريقة الثالثة فإن:

$$y_i^* = y_i - 0.380107 y_{i-1}$$

$$x_i^* = x_i - 0.380107 x_{i-1}$$

وبإستخدام طريقة المربعات الصغرى للبيانات المحولة كانت النتيجة:

$$\hat{y}^* = -1727.4 + 0.290662x^*$$

وقيمة  $DW = 1.315$  ونلاحظ أن قيمة  $DW$  تقع بين  $d_L$  و  $d_U$  أي بين 1.20 و 1.41 والتي تعني عدم وجود الارتباط الذاتي

مثال (٤-٨)

يعطى جدول (٨-٨) بيانات لمتغيرين  $Y$  ,  $x$  والمطلوب اختبار الارتباط الذاتي ومعالجته.

جدول (٨-٨)

الفترة $i$	(1) $y_i$	(2) $x_i$	(3) $e_i$	(4) $e_i e_{i-1}$	(5) $e_i^2$
1	3.63	0.97	0.2812	-	0.0791
2	4.20	0.95	0.3654	0.1028	0.1335
3	3.33	0.99	0.4670	0.1706	0.2181
4	4.54	0.91	-0.2662	-0.1243	0.0709
5	2.89	0.98	-0.2159	0.0575	0.0466
6	4.87	0.90	-0.1791	0.0387	0.0321
7	4.90	0.89	-0.3920	0.0702	0.1537
8	5.29	0.86	-0.7307	0.2864	0.5339
9	6.18	0.85	-0.0836	0.0611	0.0070
10	7.20	0.82	0.2077	-0.0174	0.0431
11	7.25	0.79	-0.4710	-0.0978	0.2218
12	6.09	0.83	-0.6594	0.3106	0.4348
13	6.80	0.81	-0.4352	0.2870	0.1894
14	8.65	0.77	0.4432	-0.1929	0.1964
15	8.43	0.76	-0.0197	-0.0087	0.0004
16	8.29	0.80	0.8119	-0.0160	0.6592
17	7.18	0.83	0.4306	0.3496	0.1854
18	7.90	0.79	0.1790	0.0771	0.0320
19	8.45	0.76	0.0003	0.0001	0.0000
20	8.23	0.78	0.2661	0.0001	0.0708

$$\sum_{i=2}^{20} e_i e_{i-1} = 1.3547 \quad \sum_{i=1}^{20} e_i^2 = 3.3082$$

**الحل**

باستخدام طريقة المربعات الصغرى فإن معادلة الانحدار المقدرة هي:

$$\hat{y} = 26.90989 - 24.28977x$$

الممود 3 في جدول (٨-٨) يوضح البواقي لهذا النموذج وعلى ذلك فإن قيمة DW هي 1.14 والتي عند مقارنتها مع القيم الحرجة عند  $\alpha = 0.05$  و  $n = 20$  حيث  $d_U = 1.41$ ,  $d_L = 1.2$  توضح أن هناك ارتباط ذاتي موجب لأن  $0 < 1.14 < d_L$ . التقدير لمعامل الارتباط الذاتي بحسب من المعادلة:

$$\hat{\rho} = \frac{\sum_{i=2}^{20} (e_i e_{i-1})}{\sum_{i=1}^{20} e_i^2} = \frac{1.3547}{3.3082} = 0.409.$$

البيانات المحولة تحسب كالتالي:

$$x_i^* = x_i - 0.409x_{i-1}, \quad y_i^* = y_i - 0.409y_{i-1}$$

حيث  $i = 1, 2, \dots, 20$ . هذه البيانات المحولة موضحة في جدول (٨-٩). معادلة الانحدار المقدرة باستخدام طريقة المربعات الصغرى سوف تكون:

$$\hat{y}^* = 15.85043 - 24.19991x^*$$

جدول (٨-٩)

i	(1) $x_i$	(2) $y_i$	(3) $e_i$
2	0.553	2.715	0.2504
3	0.601	1.612	0.3176
4	0.505	3.178	-0.4572
5	0.608	1.033	-0.1070
6	0.499	3.688	-0.0908
7	0.522	2.908	-0.3187
8	0.496	3.286	-0.5704
9	0.498	4.016	0.2153
10	0.472	4.672	0.2419
11	0.455	4.305	-0.5539
12	0.507	3.125	-0.4668
13	0.471	4.309	-0.1655
14	0.439	5.869	0.6212
15	0.445	4.892	-0.2010
16	0.489	4.842	0.8200
17	0.503	3.789	0.0985
18	0.451	4.963	0.0029
19	0.437	5.219	-0.0729
20	0.469	4.774	0.2660

قيمة الإحصاء DW للنموذج المحول هو  $DW = 1.94$ . وبمقارنة هذه القيمة عند  $\alpha = 0.05$  و  $n = 19$ ،  $d_U = 1.41$ ،  $d_L = 1.20$  (تقريباً) وبما أن  $1.41 < 1.94 < 2$  أي  $d_U < 1.94 < 2$ ، فإننا نستنتج أن الأخطاء للنموذج المحول غير مرتبطة وعلى ذلك فإن الطريقة الثالثة اختزلت مشكلة الارتباط الذاتي.



ويجب أن ننوه هنا إلى أن  $\beta_1^*$  في النموذج المحول تساوي  $\beta_1$  والموجودة في النموذج الأصلي (١-٨) وعلى ذلك بمقارنة جدول (١٠-٨) والذي يعتمد على البيانات الأصلية و جدول (١١-٨) والذي يعتمد على البيانات المحولة نجد أن هذه الطريقة الثالثة أنت الى تقدير للميل يختلف قليلا عن الذي حصلنا عليه باستخدام طريقة المربعات الصغرى . بمقارنة الأخطاء المعيارية للتقديرات من جدول (١٠-٨) و (١١-٨) نجد أن تقدير الميل من الطريقة الثالثة له خطأ معياري أكبر من تقدير المربعات الصغرى العادية. بدلالة المتغيرات الأصلية فإن الجزء المقطوع من محور الصادات وخطأ المعاري هو:

$$b_0 = \frac{b_0^*}{1-\rho} = \frac{15.85043}{1-0.409} \\ = 26.81968 ,$$

الخطأ المعياري له هو:

$$s.e(B_0) = \frac{s.e(B_0^*)}{1-\hat{\rho}} = \frac{0.9471}{1-0.409} = 1.6025.$$

جدول (١٠-٨)

المعاملات	التقدير	الخطأ المعياري
$\beta_0$	26.90989	1.1099
$\beta_1$	-24.28977	1.2978
MSE = 0.1838		$R^2 = 0.95$

جدول (١١-٨)

المعاملات	التقدير	الخطأ المعياري
$\beta_0^*$	15.85043	0.9471
$\beta_1^*$	-24.19991	1.9015
MSE = 0.1547		$R^2 = 0.91$

### (٨-٤-٤) الطريقة الرابعة

قبل تناول هذه الطريقة سوف نتناول خواص حدود الخطأ.

#### خواص حدود الخطأ

علمنا مما سبق ان الخطأ العشوائي لكل فترة زمنية يعتمد بشكل خطي على الخطأ العشوائي للفترة السابقة لها (أي أن :

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i ,$$

ومن النموذج (٨-١) يمكن الحصول على:

$$\varepsilon_{i-1} = \rho \varepsilon_{i-2} + u_{i-1} ,$$

وبالتعويض نحصل على:

$$\varepsilon_i = \rho(\rho \varepsilon_{i-2} + u_{i-1}) + u_i = \rho^2 \varepsilon_{i-2} + \rho u_{i-1} + u_i ,$$

والآن بوضع  $\varepsilon_{i-2} - \rho \varepsilon_{i-3}$  مكان  $\varepsilon_{i-2}$  نحصل على:

$$\varepsilon_i = \rho^3 \varepsilon_{i-3} + \rho^2 \varepsilon_{i-2} + \rho u_{i-1} + u_i ,$$

وبالاستمرار بهذه الطريقة نجد أن :

$$\varepsilon_i = u_i + \rho u_{i-1} + \rho^2 u_{i-2} + \rho^3 u_{i-3} + \dots = \sum_{s=0}^{\infty} \rho^s u_{i-s} , \quad (٨-١٣)$$

أي ان الخطأ في الفترة  $i$  يمثل تركيبه خطية من حد الاضطراب الراهن  $u_i$  والحدود السابقة له . وعندما  $0 < \rho < 1$  فإن (٨-١٣) تشير إلى أن كلما بعدت الفترة في الماضي كلما كان لحد الاضطراب وزن أقل في تحديد قيمة  $\varepsilon_i$  . ويمكن اثبات أن المتوسط لـ  $\varepsilon_i$  في نموذج خط الانحدار الذاتي من الرتبة الأولى (٨-٢) هي كالآتي:

$$E(\varepsilon_i) = 0$$

وذلك بأخذ توقع  $\varepsilon_i$  في (٨-١٣).

تباين الأخطاء العشوائية في حالة الارتباط الذاتي يكون كالآتي:

$$E(\varepsilon_i^2) = E(u_i^2) + \rho^2 E(u_{i-1})^2 + \rho^4 E(u_{i-2})^2 + \dots$$

وبما أن :

$$E(u_i^2) = \sigma_u^2, \quad E(u_i u_j) = 0, \quad i \neq j$$

أذن :

$$\sigma_\varepsilon^2 = \sigma_u^2 + \rho^2 \sigma_u^2 + \rho^4 \sigma_u^2 + \dots$$

أي أن :

$$\sigma_\varepsilon^2 = \sigma_u^2 (1 + \rho^2 + \rho^4 + \dots)$$

$$= \sigma_u^2 / (1 - \rho^2) \quad (٨-١٤)$$

أما التعابير بين الأقواس لمشوائية النموذج الخطي المدروس يمكن الوصول إليه بالشكل التالي:

$$\varepsilon_i = u_i + \rho u_{i-1} + \rho^2 u_{i-2} + \dots$$

وكذلك:

$$\varepsilon_{i-1} = u_{i-1} + \rho u_{i-2} + \rho^2 u_{i-3} + \dots$$

أذن:

$$E(\varepsilon_i \varepsilon_{i-1}) = E[(u_i + \rho u_{i-1} + \rho^2 u_{i-2} + \dots)$$

$$(u_{i-1} + \rho u_{i-2} + \rho^2 u_{i-3} + \dots)]$$

$$= E\{[u_i + \rho(u_{i-1} + \rho u_{i-2} + \dots)][u_{i-1} + \rho(u_{i-2} + \rho u_{i-3} + \dots)]\}$$

وعلى ذلك:

$$E(\varepsilon_i \varepsilon_{i-1}) = \rho E(u_{i-1} + \rho u_{i-2} + \dots)^2$$

$$= \rho [E(u_{i-1}^2) + \rho^2 E(u_{i-2}^2) + \dots]$$

$$= \rho [\sigma_u^2 + \rho^2 \sigma_u^2 + \dots]$$

أذن:

$$E(\varepsilon_i \varepsilon_{i-1}) = \rho \sigma_u^2 (1 + \rho^2 + \rho^4 + \dots)$$

$$= \frac{\rho \sigma_u^2}{1 - \rho^2} \quad (١٥-٨)$$

وبمقارنة (١٥-٨) مع (١٤-٨) نجد أن

$$E(\varepsilon_i \varepsilon_{i-1}) = \rho \sigma_\varepsilon^2 \quad (١٦-٨)$$

وللتسهيل سوف نضع  $\sigma_\varepsilon^2 = \sigma^2$

والعلاقة في (١٦-٨) يمكن أن توضع بشكل عام كالتالي:

$$E(\varepsilon_i \varepsilon_{i-s}) = \rho^s \sigma^2, \quad s = 0, 1, 2, \dots, n-1.$$

فعلى سبيل المثال لو كانت  $s = 0$  فإن :

$$E(\varepsilon_i \varepsilon_{i-0}) = \sigma^2$$

وفي حالة  $s=1$  فإن:

$$E(\varepsilon_i \varepsilon_{i-1}) = \rho \sigma^2$$

أما إذا كانت  $s = 2$  فإن :

$$E(\varepsilon_i \varepsilon_{i-2}) = \rho^2 \sigma^2$$

وأخيراً إذا كانت  $s = n-1$  فإن:

$$E(\varepsilon_i \varepsilon_{i-(n-1)}) = \rho^{n-1} \sigma^2$$

وعلى ذلك معامل الارتباط بين حدود الأخطاء هو:

$$\rho_{\varepsilon_i \varepsilon_{i-1}} = \frac{\text{Cov}(\varepsilon_i, \varepsilon_{i-1})}{\sigma_{\varepsilon_i} \sigma_{\varepsilon_{i-1}}}$$

$$= \frac{\rho \left( \frac{\sigma^2}{1 - \rho^2} \right)}{\sqrt{\frac{\sigma^2}{1 - \rho^2}} \sqrt{\frac{\sigma^2}{1 - \rho^2}}} = \rho.$$

أي أن معلمة الارتباط الذاتي  $\rho$  هي نفسها معامل الارتباط بين حدود الأخطاء المتجاورة حيث  $\sigma_u^2 = \sigma^2$ .

ويجمع هذه الحدود في مصفوفة التغاير والتباين للأخطاء العشوائية في حالة النموذج الخطي المتعدد الذي على الصورة التالية :

$$Y = X\beta + \varepsilon \quad (١٧-٨)$$

نحصل على :

$$\text{Cov}(\varepsilon\varepsilon') = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & \dots & \rho^{n-1}\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \dots & \rho^{n-2}\sigma^2 \\ . & . & . & . & . \\ \rho^{n-3}\sigma^2 & \rho^{n-2}\sigma^2 & \rho^{n-3}\sigma^2 & \dots & \sigma^2 \end{bmatrix}$$

أي أن :

$$\text{Cov}(\varepsilon\varepsilon') = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \dots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 \dots & \rho^{n-2} \\ . & . & . & . & . \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} \dots & 1 \end{bmatrix} = \sigma^2 \Sigma$$

أي أن الخطأ العشوائي لنموذج الإحداد الخطي المتعدد (١٧-٨) سوف يخضع لفرضية وجود ارتباط ذاتي من الدرجة الأولى - أي أن :

$$\varepsilon \sim N(0, \sigma^2 \Sigma)$$

ولتقدير معالم نموذج الإحداد الخطي (١٧-٨) سوف نتبع الطريقة الرابعة (طريقة المربعات الصغرى المرجحة) والتي تختلف عما أوضحناه في الفصل الرابع فقط من حيث أن المصفوفة  $\Sigma$  ليست قطرية . وفي هذه الطريقة لابد من إيجاد معكوس المصفوفة  $\Sigma$  ويجب أن تكون رتبة المصفوفة  $\Sigma$  مساوية إلى حجم العينة تحت البحث . فمثلا عند العينة  $n=2$  تكون المصفوفة  $\Sigma$  على الشكل التالي :

... .

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

أما إذا كان حجم العينة  $n=3$  فإن :

$$\Sigma = \begin{bmatrix} 1 & 1+\rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}$$

وعلى ذلك فإن معكوس المصفوفة  $\Sigma$  سوف يكون:

$$\Sigma^{-1} = \frac{1}{(1-\rho^2)} \begin{bmatrix} 1 & -\rho & 0 \\ -\rho & 1+\rho^2 & -\rho \\ 0 & -\rho & 1 \end{bmatrix}$$

وبصورة عامة فإن المصفوفة  $\Sigma^{-1}$  لعينة من الحجم  $n$  سوف تكون :

$$W = \Sigma^{-1} = \frac{1}{(1-\rho^2)} \begin{bmatrix} 1 & -\rho & 0 & 0 & 0 \dots 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 \dots 0 & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 \dots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & 0 & 0 & -\rho & 1 \end{bmatrix}$$

التقدير لمصفوفة التغاير والتباين سوف تكون :  $s^2(X'\Sigma^{-1}X)^{-1}$

حيث أن :

$$s^2 = \frac{y'\Sigma^{-1}y - b'X'\Sigma^{-1}y}{n-k-1}$$

الآن سوف نشرح الطريقة الرابعة لتقدير معالم نموذج الانحدار (٨-١٧) والمسمى بطريقة للمربعات الصغرى المرجحة وذلك من خلال المثال التالي .

مثال (٨-٥)

عينة عشوائية ذات خمسة مشاهدات ، اخذ فيها كل من المتغير التابع (Y) والمتغيرات المستقلة  $(x_1), (x_2)$  المشاهدات التالية:

$$y: 4, 8, 6, 2, 9$$

$$x_1: 2, 5, 2, 1, 10$$

$$x_2: 1, 3, 7, 2, 1$$

المطلوب :

تقدير معالم النموذج التالي:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

مستخدماً:

الطريقة الرابعة علماً بأن:

$$\varepsilon \sim N(0, \sigma^2 \Sigma)$$

وأن  $(\varepsilon)$  يتبع الارتباط الذاتي من الدرجة الاولى مع  $(\hat{\rho} = 0.6)$

الحل

أولاً:

$$b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = (X'WX)^{-1} X'Wy$$

علماً بأن:

$$W = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 \\ 0 & 0 & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & -\rho & 1 \end{bmatrix}$$

لأن:

$$W = \frac{1}{0.64} \begin{bmatrix} 1 & -0.6 & 0 & 0 & 0 \\ -0.6 & 1.36 & -0.6 & 0 & 0 \\ 0 & -0.6 & 1.36 & -0.6 & 0 \\ 0 & 0 & -0.6 & 1.36 & -0.6 \\ 0 & 0 & 0 & -0.6 & 1 \end{bmatrix}$$

لأن :

$$(X'WX)^{-1} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 5 & 2 & 1 & 10 \\ 1 & 3 & 7 & 2 & 1 \end{bmatrix} \left( \frac{1}{0.64} \right)$$



$$\begin{bmatrix} 1 & -0.6 & 0 & 0 & 0 \\ -0.6 & 1.36 & -0.6 & 0 & 0 \\ 0 & -0.6 & 1.36 & 0.6 & 0 \\ 0 & 0 & -0.6 & 1.36 & -0.6 \\ 0 & 0 & 0 & -0.6 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 5 & 3 \\ 1 & 2 & 7 \\ 1 & 1 & 2 \\ 1 & 10 & 1 \end{bmatrix}^{-1}$$

$$= 0.64 \begin{bmatrix} 1.28 & 6.08 & 2.72 \\ 6.08 & 106.4 & 3.76 \\ 2.72 & 3.76 & 38.32 \end{bmatrix}^{-1}$$

اما:

$$X'Wy = \frac{1}{0.64} \begin{bmatrix} 7.76 \\ 98.84 \\ 24.2 \end{bmatrix},$$

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = 0.64 \begin{bmatrix} 1.28 & 6.08 & 2.72 \\ 6.08 & 106.4 & 3.76 \\ 2.72 & 3.76 & 38.32 \end{bmatrix}^{-1} \begin{bmatrix} 7.76 \\ 98.84 \\ 24.2 \end{bmatrix}$$

ای ان:

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.981 \\ 0.856 \\ 0.478 \end{bmatrix}$$

آن:

$$\hat{y} = 0.981 + 0.856x_1 + 0.478x_2.$$

## الفصل التاسع

### الارتباط الخطي المتعدد

### Multicollinearity

- (١-٩) مقدمة
- (٢-٩) مصادر الارتباط الخطي المتعدد
- (٣-٩) تأثيرات الارتباط الخطي المتعدد
- (٤-٩) مؤشرات لوجود الارتباط الخطي المتعدد
- (٥-٩) طرق الكشف عن الارتباط الخطي المتعدد
- (١-٥-٩) فحص مصفوفة الارتباط
- (٢-٥-٩) عوامل تضخم التباين
- (٣-٥-٩) تحليل القيم المميزة
- (٤-٥-٩) تشخيصات أخرى
- (٦-٩) معالجة الارتباط الخطي المتعدد
- (٧-٩) التحذار المكونات الرئيسية

## (١-٩) مقدمة

تميل المتغيرات المستقلة في العديد من الدراسات في مجال الأعمال، الاقتصاد، والعلوم الاجتماعية والبيولوجية، إلى أن تكون مرتبطة فيما بينها ومرتبطة مع متغيرات أخرى ذات صلة بالمتغير التابع وغير موجوده في النموذج. على سبيل المثال، في انحدار نفقات الطعام للأسره على المتغيرات المستقلة: دخل الأسرة، توفيرات الأسرة، وعمر رب الأسرة، ستكون المتغيرات المستقلة مرتبطة فيما بينها. وأكثر من ذلك ستكون المتغيرات المستقلة مرتبطة أيضا بمتغيرات اجتماعية - اقتصادية غير موجوده في النموذج ولها تأثيرها على نفقات طعام الأسرة، مثل حجم الامرء. وعندما تكون المتغيرات المستقلة مرتبطة فيما بينها يقال أنه يوجد ارتباط خطي متعدد فيما بينها.

سوف نكتب نموذج الانحدار المتعدد على الصورة التالية:

$$Y = X\beta + \varepsilon \quad (1-9)$$

حيث  $Y$  متجه من الدرجة  $n \times 1$  من الاستجابات و  $X$  مصفوفة من الدرجة  $n \times k$  من المتغيرات المستقلة و  $\beta$  متجه من الدرجة  $k \times 1$  من الثوابت الغير معلومة و  $\varepsilon$  متجه من الدرجة  $n \times 1$  من الأخطاء العشوائية المستقلة حيث  $(\varepsilon_j \sim NID(0, \sigma^2))$ . سوف نفترض أن المتغيرات المستقلة والاستجابة في صورة قيم معيارية كما تناولنا في البند (٣-١٣) وعلى ذلك  $X'X$  مصفوفة من الدرجة  $k \times k$  من الارتباطات بين المتغيرات المستقلة و  $X'y$  متجه من الدرجة  $k \times 1$  من الارتباطات بين المتغيرات المستقلة والاستجابة. ليكن العمود رقم  $i$  من المصفوفة  $X$  والذي نرمز له بالرمز  $X_i$  وعلى ذلك  $X = [X_1, X_2, \dots, X_k]$ . وعلى ذلك  $X_i$  يحتوى على  $n$  من مستويات المتغير المستقل رقم  $i$ . سوف نعرف الارتباط الخطي المتعدد التام بدلالة عدم الاستقلال الخطي بين اعمدة  $X$ . المتجهات  $X_1, X_2, \dots, X_k$  تمثل علاقة خطية غير مستقلة إذا وجدت فئة من الثوابت  $t_1, t_2, \dots, t_k$  لا تساوي جميعها الصفر حيث:

$$\sum_{i=1}^k t_i X_i = 0 \quad (2-9)$$

ومن المعادلة (٢-٩) يمكن اشتقاق اي متغير مستقل  $k, i = 1, 2, \dots$  كتركيبية خطية من بقية المتغيرات المستقلة على النحو التالي:

$$X_1 = -\frac{t_2 X_2}{t_1} - \frac{t_3 X_3}{t_1} - \dots - \frac{t_k X_k}{t_1}, \quad t_1 \neq 0$$

$$X_2 = \frac{-t_1 X_1}{t_2} - \frac{t_3 X_3}{t_2} - \dots - \frac{t_k X_k}{t_2}, \quad t_2 \neq 0$$

وهكذا لبقية المتغيرات المستقلة.

عندما نتحقق (٩-٢) بالضبط لفئة جزئية من أعمدة  $X$ ، فإن رتبة المصفوفة  $X'X$  تكون أقل من  $k$  وتكون المصفوفة  $(X'X)$  مصفوفة شاذة أي أن محددها يساوي الصفر، وبالتالي فإن معكوس  $X'X$  يكون غير موجود. وعلى العكس إذا لم يكن بين المتغيرات المستقلة أي علاقة خطية أي كان معامل الارتباط بينها مساويا للصفر، سميت هذه المتغيرات بالمتعامدة orthogonal أي المتغيرات التي يكون تغايرها مساويا للصفر. وفي أغلب الأحيان نجد أن بين المتغيرات المستقلة درجة من الارتباط.

#### (٩-٢) مصادر الارتباط الخطي المتعدد

١. ميل بعض المتغيرات للتحرك معا مع مرور الزمن، فعلى سبيل المثال دخل الموظف وسنوات خبرته وعمره ومرتبته الوظيفية غالبا ما تتغير سويا ويكون بينهما ارتباط خطي قوي.

٢. استخدام بعض المتغيرات المستقلة بفترات تأخير ومن الطبيعي أن القيم المتعاقبة لمتغير معين يكون بينها علاقة فالدخل في الفترة الحالية يتحدد جزئيا عن طريق قيمته في الفترة السابقة وهكذا. ولذا فإن مشكلة الارتباط الخطي غالبا ماتكون موجودة مؤكدا في نماذج فترات التأخير.

٣. قلة عدد المشاهدات مقارنة بعدد المتغيرات الموجودة في النموذج وهذه الحالة تحدث في الأبحاث الطبية والإنسانية حيث عدد الأشخاص تحسب الدراسة قليل والمعلومات تجمع على عدد كبير من المتغيرات المستقلة تحت الدراسة. الأسلوب المفيد في هذه الحالة هو حذف بعض المتغيرات المستقلة من الدراسة. لقد قدم (Mason et al 1975) وآخرون ثلاثة توصيات.

- أ- اعاده توصيف للنموذج بدلالة عدد صغير من المتغيرات المستقلة.
- ب- إجراء بحث مبدئي باستخدام فئات جزئية من المتغيرات المستقلة الأصلية.
- ج- استخدام تحليل المكونات الأساسية والذي سوف نتناوله في البند (٩-٧).

### (٣-٩) تأثيرات الارتباط الخطي المتعدد

إن وجود الارتباط الخطي له تأثيرات خطيرة على تقديرات المربعات الصغرى لمعاملات الانحدار. يفرض وجود متغيرين مستقلين  $x_1, x_2$  في النموذج (٣-٩) ويفرض أن قيم  $x, Y$  معيارية، نموذج الانحدار سوف يكون:

$$Y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

المعادلات الطبيعية للمربعات الصغرى سوف تكون:

$$(X'X)b = X'y.$$

أي أن:

$$\begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} r_{y1} \\ r_{y2} \end{bmatrix}$$

حيث  $r_{12}$  هو معامل الارتباط البسيط بين  $x_1, x_2$  و  $r_{21} = r_{12}$  ، و  $r_{yi}$  هو معامل الارتباط البسيط بين  $Y, x_i$  ،  $i = 1, 2$ . الآن معكوس  $X'X$  سيكون:

$$C = (X'X)^{-1} = \begin{bmatrix} \frac{1}{(1-r_{12}^2)} & \frac{-r_{12}}{(1-r_{12}^2)} \\ \frac{-r_{12}}{(1-r_{12}^2)} & \frac{1}{(1-r_{12}^2)} \end{bmatrix} \quad (٣-٩)$$

وتقديرات معاملات الانحدار سوف تكون:

$$b_1 = \frac{r_{y1} - r_{12}r_{y2}}{(1-r_{12}^2)}, \quad b_2 = \frac{r_{y2} - r_{12}r_{y1}}{(1-r_{12}^2)}. \quad (٤-٩)$$

عند وجود ارتباط خطي قوي بين  $x_1, x_2$  فإن معامل الارتباط  $r_{12}$  سوف يكون كبير. من (٣-٩) سوف نجد أن:

$$\text{Cov}(B_1, B_2) = c_{12}\sigma^2 \rightarrow \pm\infty, \quad \text{Var}(B_i) = c_{ii}\sigma^2 \rightarrow \infty, \quad |r_{12}| \rightarrow 1,$$

وذلك بالاعتماد على أن  $r_{12} \rightarrow +1$  أو  $r_{12} \rightarrow -1$ . هذا الارتباط الخطي القوي بين  $x_1, x_2$  يؤدي إلى تباينات وتغايرات كبيرة لمقدرات المربعات الصغرى لمعاملات الانحدار. وهذا يؤدي إلى أن العينات المختلفة ولماخوذة عدد نفس المستويات من  $x$  قد تعطي تقديرات مختلفة بدرجة كبيرة لمعامل النموذج. عند وجود أكثر من متغيرين مستقلين، فإن مشكلة الارتباط الخطي تعطي نفس التأثير. ويمكن إثبات أن العناصر على القطر للمصفوفة  $C = (X'X)^{-1}$  هم:

$$c_{ii} = \frac{1}{1 - R_i^2}, \quad i = 1, 2, \dots, k.$$

حيث  $R_i^2$  هو معامل التحديد المتعدد عند بناء نموذج انحدار للمتغير  $x_i$  على بقية المتغيرات المستقلة. عندما يوجد علاقة خطية قوية بين  $x_i$  وأي فئة جزئية من المتغيرات المستقلة الأخرى التي عددها  $k-1$  فإن القيمة  $R_i^2$  سوف تقترب من الواحد الصحيح. وبما أن تبعاين  $B_i$  هو  $\text{Var}(B_i) = c_{ii}\sigma^2 = (1 - R_i^2)^{-1}\sigma^2$  فإن علاقة خطية قوي تؤدي إلى أن تباین مقدرات المربعات الصغرى لمعاملات الانحدار  $\beta_i$  سوف تكون كبيرة جدا. عموما التباينات لـ  $B_i, B_{i'}, B_i \neq i'$  سوف تكون كبيرة إذا كان المتغيران  $x_i, x_{i'}$  يشملان في العلاقة الخطية. أيضا يؤدي الارتباط الخطي إلى أن تقديرات المربعات الصغرى  $b_i$  تكون كبيرة جدا في القيمة المطلقة. لاثبات ذلك بفرض أن المسافة المربعة بين  $B$  إلى متجه المعلمة الحقيقية هو:

$$L_1^2 = (B - \beta)'(B - \beta). \quad (٥-٩)$$

القيمة المتوقعة لـ  $L_1^2$  هي:

$$\begin{aligned} E(L_1^2) &= E(B - \beta)'(B - \beta) \\ &= \sum_{i=1}^k E(B_i - \beta_i)^2 \\ &= \sum_{i=1}^k \text{Var}(B_i) \\ &= \sigma^2 \text{tr}(X'X)^{-1}. \end{aligned} \quad (٦-٩)$$

حيث  $\text{tr}$  هو أثر المصفوفة والذي يساوي حاصل جمع عناصر القطر الرئيسي للمصفوفة. عند وجود مشكلة الارتباط الخطي المتعدد فإن بعض القيم المميزة للمصفوفة  $X'X$  سوف تكون صغيرة. وبما أن أثر المصفوفة أيضا يساوي حاصل جمع القيم المميزة للمصفوفة  $X'X$  فإن (٦-٩) تصبح:

$$E(L_1^2) = \sigma^2 \sum_{i=1}^k \frac{1}{\lambda_i}, \quad (٧-٩)$$

حيث  $\lambda_i > 0$ ,  $i = 1, 2, \dots, k$  هم القيم المميزة للمصفوفة  $X'X$ . لن وجود واحد على الأقل من القيم المميزة صغير في (٧-٩) يؤدي إلى أن المسافة بين مقدر المربعات الصغرى  $B$  إلى المعلمة الحقيقية  $\beta$  كبيرة. وهذا يكافئ أن:

$$\begin{aligned} E(L_1^2) &= E(B - \beta)'(B - \beta) \\ &= E(B'B - 2B'\beta + \beta'\beta) \end{aligned}$$

أو:

$$E(B'B) = \beta'\beta + \sigma^2 \text{tr}(X'X)^{-1}$$

أي أن المتجه  $B$  عادة يكون أطول من المتجه  $\beta$ . وهذا يعني أن طريقة المربعات الصغرى تؤدي إلى تقديرات لمعاملات الانحدار كبيرة في القيمة المطلقة. وعلى ذلك فإن تقديرات المربعات الصغرى سوف تعاني من عدم الدقة.

#### (٩-٤) مؤشرات لوجود الارتباط الخطي المتعدد

١. تغيرات كبيرة في معاملات الانحدار المقدرة عند إضافة أو حذف متغير أو عند تعديل أو حذف مشاهدة.

٢. نتائج غير معنوية لاختبارات فردية حول معاملات الانحدار الخاصة بمتغيرات مستقلة مهمة.

٣. معاملات انحدار مقدرة، اشارتها للجبرية معاكسة تماماً لما تتوقعه الاعتبارات النظرية أو الخبرة السابقة.

٤. معاملات كبيرة للارتباط البسيط بين أزواج المتغيرات المستقلة في مصفوفة الارتباط.

٥. فترات ثقة عريضة لمعاملات انحدار تمثل متغيرات مستقلة مهمة.

مثال (٩-١)

يعطي الجدول (٩-١) بيانات تم توليدها على الحاسب الآلي بحيث يوجد ارتباط خطي بين  $x_1$ ,  $x_2$ . كما أن  $y^{(1)}$ ,  $y^{(2)}$ ,  $y^{(3)}$  تمثل عينات مختلفة.

## جدول (٩-١)

$x_1$	$x_2$	$y^{(1)}$	$y^{(2)}$	$y^{(3)}$
2.705	2.695	4.10	4.10	4.06
2.995	3.005	4.34	4.73	4.39
3.255	3.245	4.95	4.81	5.02
3.595	3.605	5.36	5.30	5.23
3.805	3.795	5.64	5.75	5.57
4.145	4.155	6.18	6.26	6.50
4.405	4.395	6.69	6.61	6.65
4.745	4.755	7.24	7.13	7.26
4.905	4.895	7.46	7.30	7.48
4.845	4.855	7.23	7.32	7.39

وبفرض النموذج:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$$

فإن تقديرات المربعات الصغرى للمتجه  $b$  للمتغير التابع  $Y^{(1)}$  هي:

$$b_0 = -0.209(-1.634), b_1 = 5.897(2.448), b_2 = -4.341(-1.806)$$

حيث القيم بين الاقواس تمثل قيم  $t$ . للمتغيران  $Y^{(2)}, Y^{(3)}$  فإن التقديرات المقابلة هي:

$$b_0 = 0.182(1.158), b_1 = -2.002(-0.676), b_2 = 3.461(1.171),$$

$$b_0 = -0.298(-1.223), b_1 = 1.526(0.332), b_2 = 0.06109(0.013).$$

حيث يلاحظ من النتائج السابقة اختلاف كبير في قيم  $b_i$  المقدره.

**(٩-٥) طرق الكشف عن الارتباط الخطي المتعدد**

هناك أساليب عديدة للكشف عن الارتباط الخطي المتعدد. في هذا البند سوف نناقش ونبسط بعض المقاييس للكشف عن الارتباط الخطي المتعدد. بعض هذه المقاييس تمتاز بالكشف المباشر عن درجة الارتباط الخطي المتعدد وامدادنا بمعلومات تساعدنا في تقدير أي من المتغيرات المستقلة سبب هذه المشكلة.



### (٩-٥-١) فحص مصفوفة الارتباط

يعتبر فحص مصفوفة الارتباط أبسط مقياس للكشف عن الارتباط الخطي المتعدد حيث يتم فحص معاملات الارتباط الخطي البسيط بين أزواج المتغيرات المستقلة  $i \neq i', i_{jj'}$  والتي تقع فوق القطر الرئيسي في المصفوفة  $X'X$  (المصفوفة المعتمدة على القيم المعيارية لكل من قيم  $x$  و  $y$ ) وبملاحظة قيم معاملات الارتباط إذا وجد أن هناك ارتباط قويا بين أي متغيرين مستقلين دل ذلك على احتمال وجود ارتباط خطي متعدد.

ويجب ملاحظة أن ضعف العلاقة الزوجية بين المتغيرات المستقلة لا يعني غياب المشكلة إذ يمكن أن يكون هناك علاقة خطية أو تركيب خطي بين أحد المتغيرات المستقلة ومتغيرين أو أكثر من بقية المتغيرات المستقلة.

### مثال (٩-٢)

يعطي جدول (٩-٢) بيانات عن الاتفاق على الملابس والدخل التصرفي والاصول المسئلة والرقم القياسي لاسعار الملابس والرقم القياسي العام لاسعار خلال الفترة من 59 إلى 68. والمطلوب فحص مصفوفة الارتباط للتعرف على وجود أو عدم وجود ارتباط خطي.

جدول (٩-٢)

الرقم القياسي للإسعار ( $x_2$ ) ١٠٠=١٩٦٣	الرقم القياسي لاسعار الملابس ( $x_3$ ) ١٠٠=١٩٦٣	الاصول المستقلة ( $x_3$ )	الدخل التصرفي ( $x_1$ )	الاتفاق ( $y$ )	السنة
94	92	17.1	82.9	8.4	1959
96	93	21.3	88.0	9.6	1960
97	96	25.1	99.9	10.4	1961
97	94	29.0	105.3	11.4	1962
100	100	34.0	117.7	12.2	1963
101	101	40.0	131.0	14.2	1964
104	105	44.0	148.2	15.8	1965
109	112	49.0	161.8	17.9	1966
111	112	51.0	174.2	19.3	1967
111	112	53.0	184.7	20.8	1968

### الحل

مصفوفة الارتباط للبيانات المعطاه في جدول (٩-٢) هي:

$$X'X = \begin{bmatrix} 1 & 0.980 & 0.988 & 0.988 \\ & 1 & 0.970 & 0.992 \\ & & 1 & 0.969 \\ & & & 1 \end{bmatrix}$$

حيث  $X'X$  تعتمد على القيم المعيارية لقيم المتغيرات المستقلة تعكس المصفوفة  $X'X$  ارتباط قوي بين  $x_1$ ,  $x_3$  وذلك لان  $r_{13} = 0.988$ . ونفس الشيء بين  $x_1$ ,  $x_4$  كما أن الارتباطات الأخرى عالية. أي يوجد ارتباطات قوية بين المتغيرات وبعضها.

### (٩-٥-٢) عوامل تضخم التباين

تسمى العناصر القطرية في المصفوفة  $(X'X)^{-1}$  (حيث المصفوفة  $X'X$  على شكل مصفوفة الارتباط) عوامل تضخم التباين ( $VIF_i$ ) حيث يمكن اعتبارهم مقياس هام للكشف عن الارتباط الخطي. مره أخرى من (٩-٣) فإن  $c_{ii}$  العنصر رقم  $i$  على القطر للمصفوفة  $C$  يمكن كتابته على الشكل  $c_{ii} = (1 - R_i^2)^{-1}$  ، حيث  $R_i^2$  هو معامل التحديد الذي نحصل عليه من نموذج انحدار المتغير المستقل رقم  $i$  على بقية المتغيرات المستقلة وعددها  $k-1$ . عندما  $x_i$  يكون قريب من الاعتماد على المتغيرات المستقلة الباقية فإن  $R_i^2$  سوف يكون صفر و  $c_{ii}$  يقترب من الواحد الصحيح ، بينما إذا كان  $x_i$  على علاقة شبة خطيه مع بعض الفئات الجزئية من المتغيرات فإن  $R_i^2$  تقترب من الواحد الصحيح و  $c_{ii}$  تصبح كبيرة. بما أن التباين لمعامل الانحدار رقم  $i$  هو  $c_{ii}\sigma^2$  فإننا يمكن النظر إلى  $c_{ii}$  كمعامل يؤدي إلى تضخم تباينات مقدرات المربعات الصغرى  $B_i$  نتيجة للارتباط الخطي القوي. يمكن تعريف معامل تضخم التباين كالآتي:

$$VIF_i = c_{ii} = (1 - R_i^2)^{-1}$$

هذا التعريف راجع إلى (1970) Marquardt. كبر واحد أو أكثر من  $VIF$  يدل على وجود الارتباط الخطي. نكل الخبرة التجريبية على أن أي واحد من  $VIF$  يزيد عن 10 يكون مؤشر على أن معاملات الانحدار تقديرها غير دقيق بسبب وجود الارتباط الخطي. يأخذ عامل تضخم التباين قوماً غير مساليه أي أن

$VIF_i \geq 0$ . وفي حالة ارتباط خطي تام بين المتغير المستقل  $x_i$  وبقيّة المتغيرات المستقلة فإن  $R_i^2 = 1$  وبالتالي فإن عامل تضخم التباين يتخذ قيما لانهاية وفي حالة عدم وجود ارتباط خطي بين المتغير المستقل  $x_i$  وبقيّة المتغيرات المستقلة (حالة التعمد) ، فإن  $R_i^2 = 0$  وقيمة  $VIF$  تساوي الواحد الصحيح .

أيضا فإن  $VIF$  له تفسير آخر حيث يقيس مدى كبر فترة الثقة لمعامل الانحدار رقم  $i$ . أن طول فترة الثقة لمعامل الانحدار رقم  $i$  يمكن كتابته على الصورة التالية:

$$L_i = 2(c_{ii}\hat{\sigma}^2)^{\frac{1}{2}} / t_{\alpha/2}(n-k-1)$$

وطول الفترة المقابلة والتي تعتمد على تصميم متعامد والتي تعتمد على نفس الحجم من العينة و على  $(\sum(x_{ij} - \bar{x}_i)^2)$  هي:

$$L^* = 2\delta t_{\alpha/2}(n-k-1)$$

النسبة  $L_i / L^* = (c_{ii})^{\frac{1}{2}}$ . وعلى ذلك الجذر التربيعي لـ  $VIF$  رقم  $i$  يوضح مدى كبر فترة الثقة لمعامل الانحدار رقم  $i$  بسبب وجود الارتباط الخطي.

أيضا تستخدم عوامل تضخم التباين لقياس مدى بعد مقدرات المربعات الصغرى عن قيمتها الحقيقية. حيث تأخذ القيمة المتوقعة لمجموع مربعات الفروق بين مقدرات معاملات الانحدار وقيمتها الحقيقية الصيغة التالية:

$$E(L_i^2) = E \sum_{i=1}^k (B_i - \beta_i)^2 = \sigma^2 \sum VIF_i \quad (٨-٩)$$

وكما أشرنا من قبل فإنه في حالة عدم وجود ارتباط خطي يكون قيم عوامل تضخم التباين بين المتغيرات جميعا مساويا للواحد الصحيح وفي هذه الحالة فإن (٨-٩) تأخذ الصيغة التالية:

$$E \sum_{i=1}^k (B_i - \beta_i)^2 = \sigma^2 \sum_{i=1}^k VIF_i = \sigma^2 k.$$

ومن ثم يمكن حساب النسبة التالية:

$$\frac{\sigma^2 \sum_{i=1}^k VIF_i}{k\sigma^2} = \frac{\sum_{i=1}^k VIF_i}{k}.$$

وبلاحظ أن النسبة المسابقة عبارة عن متوسط قيم عوامل تضخم التباين لمعاملات الانحدار. إذا كانت المتغيرات المستقلة متعامدة لا يوجد بينها ارتباط خطي

فإن هذه النسبة تساوي الواحد الصحيح. ولذلك نجد أنه كلما زادت قيمة متوسط عوامل تباین التضخم عن الواحد الصحيح دل على ذلك وجود الارتباط الخطي بين المتغيرات المستقلة. وتوجد بعض البرامج الجاهزة الخاصة بالانحدار والتي تعطي معكوس عامل تضخم التباين ويعرف هذا المقياس بالتحمل (Tolerance) ويتم حسابه من الصيغة التالية:

$$\text{Tolerance} = 1/VIF_i = 1 - R_i^2.$$

ولقيم التحمل التي تستخدم بواسطة هذه البرامج كحد أدنى لدخول أى متغير في النموذج هي 0.0001, 0.001, 0.01.

#### مثال (٣-٩)

يعطي جدول (٣-٩) بيانات خاصه بستة متغيرات مستقلة ومتغير استجابة.

جدول (٣-٩)

المشاهدة j	$Y_j$	$X_{1j}$	$X_{2j}$	$X_{3j}$	$X_{4j}$	$X_{5j}$	$X_{6j}$
1	10.006	8.000	1.000	1.000	1.000	0.541	-0.099
2	9.737	8.000	1.000	1.000	0.000	0.130	0.070
3	15.087	8.000	1.000	1.000	0.000	2.116	0.115
4	8.422	0.000	0.000	9.000	1.000	-2.397	0.252
5	8.625	0.000	0.000	9.000	1.000	-0.046	0.017
6	16.289	0.000	0.000	9.000	1.000	0.365	1.504
7	5.958	2.000	7.000	0.000	1.000	1.996	-0.865
8	9.313	2.000	7.000	0.0000	1.000	0.228	-0.055
9	12.960	2.000	7.000	0.000	1.0000	0.380	0.502
10	5.541	0.000	0.000	0.000	10.000	-0.798	-0.399
11	8.756	0.000	0.000	0.000	10.000	0.257	0.101
12	10.937	0.000	0.000	0.000	10.000	0.440	0.432

قيم  $VIF_i$  الخاصه بالبيانات المعطاه في جدول (٣-٩) معطاه في جدول (٤-٩).

جدول (٤-٩)

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$
181.83	161.4	265.49	297.11	1.74	1.44

من جدول (٩-٤) يتضح أن القيمة العظمى لـ VIF هي 297.14 والخاصة بالمتغير المستقل رقم 4. واضح وجود ارتباط خطي. مره أخرى فإن VIF المقابلة للمتغيرات الداخلة في الارتباط الخطي كبيرة جدا عن المرتبطة بالمتغيرين  $X_5$  ,  $X_6$ .

### (٩-٥-٣) تحليل القيم المميزة

يمكن استخدام القيم المميزة للمصفوفة  $X'X$  ,  $\lambda_1, \lambda_2, \dots, \lambda_k$  ، كمقياس لوجود الارتباط الخطي في البيانات . عند وجود واحد أو أكثر من المتغيرات بينهما ارتباط خطي قوي فإن واحد أو أكثر من القيم المميزة سوف تكون صغيرة. بعض المحللين يفضلون اختبار رقم الحالة condition number للمصفوفة  $X'X$  والمعرف كالتالي:

$$w = \frac{\lambda \max}{\lambda \min} \quad (٩-٩)$$

عوماً إذا كان رقم الحالة أقل من 100 فهذا يدل على عدم وجود مشكلة الارتباط الخطي. ارقام الحالة بين 100 , 1000 تدل على ارتباط خطي قوى وعندما تزيد  $w$  عن 1000 فهذا يدل على وجود ارتباط خطي قوى جدا. وهناك مقياس آخر يسمى مؤشر الحالة condition index والذي يتم حسابه كمايلي:

$$w_i^* = \frac{\lambda \max}{\lambda_i} \quad (٩-١٠)$$

من الواضح أن اكبر مؤشر للحالة هو رقم الحالة المعرف في (٩-٩). المؤشرات التي قيمتها أكبر من 100 مفيدة في اكتشاف وجود الارتباط الخطي. القيم المميزة للمصفوفة  $X'X$  للبيانات المعطاء في جدول (٩-٣) والخاصه بالمثال (٩-٣) هي:

$$\lambda_1 = 2.4288 , \lambda_2 = 1.5462,$$

$$\lambda_3 = 0.9221 , \lambda_4 = 0.7940,$$

$$\lambda_5 = 0.3079 , \lambda_6 = 0.0011.$$

القيم الصغيرة من القيم المميزة تدل على وجود ارتباط خطي. رقم الحالة هو:

$$w = \frac{\lambda \max}{\lambda \min} = \frac{2.4288}{0.0011} = 2208.$$

ويدل على ارتباط خطي قوي. واحد فقط من مؤشر الحالة يزيد عن 1000 وعلى ذلك يمكن استنتاج وجود علاقة خطية واحدة في البيانات.  
إن تحليل القيم المميزة يفيد في تعريف طبيعة العلاقة الخطية في البيانات.  
المصفوفة  $X'X$  يمكن كتابتها على الصورة التالية:

$$X'X = T \wedge T'$$

حيث  $\wedge$  مصفوفة قطرية بدرجة  $k \times k$ . حيث عناصرها القطرية هم القيم المميزة  $k, \lambda_i, i = 1, 2, \dots, k$ ,  $T$  مصفوفة متعامدة بدرجة  $k \times k$  عناصرها هم متجهات القيم المميزة لـ  $X'X$ . ليكن الأعمدة للمصفوفة  $T$  هم  $t_1, t_2, \dots, t_k$ . عندما تكون القيم المميزة  $\lambda_i$  قريبة من الصفر، فهذا يدل على وجود علاقة خطية قوية في البيانات، فإن العناصر المرتبطة بمتجهات القيم المميزة  $t_i$  تصف طبيعة هذه العلاقة الخطية، حيث العناصر للمتجه  $t_i$  هم المعاملات  $t_{i1}, t_{i2}, \dots, t_{ik}$  في (٢-٩). يعطى جدول (٥-٩) متجهات القيم المميزة للبيانات المعطاه في جدول (٣-٩) والخاصة بالمثال (٣-٩). أقل قيمة مميزة هي  $\lambda_6 = 0.0011$  وعلى ذلك العناصر للمتجه  $t_6$  هم معاملات الانحدار في (٢-٩).

وهذا يعني أن:

$$\begin{aligned} & -0.44768x_1 - 0.42114x_2 - 0.54169x_3 \\ & -0.57337x_4 - 0.00605x_5 - 0.00217x_6 = 0. \end{aligned}$$

ويفرض أن  $-0.00605, -0.00217$  تقريبا يساويان الصفر وبإعادة ترتيب الحدود نحصل على:

$$x_1 \approx -0.941x_2 - 1.210x_3 - 1.281x_4$$

وعلى ذلك العناصر في  $t_6$  تعكس العلاقة المستخدمة في إيجاد  $x_1, x_2, x_3, x_4$ .

جدول (٩-٥)

$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$
-39072	-33968	.67980	.07990	-.25104	-.44768
-.45560	-.05392	-.70013	.05769	-.34447	-.42114
.48264	-.45333	-.16078	.19103	.45364	-.54169
.18766	.73547	.13587	-.27645	.01521	-.57337
-.49773	-.09714	-.03185	-.56356	.65128	-.00605
.35195	-.35476	-.04864	-.74818	-.43375	-.00217

(٩-٥-٤) تشخيصات أخرى

هناك العديد من الطرق المفيدة في تشخيص الارتباط الخطي. المحدد للمصفوفة  $X'X$  يمكن استخدامه كدليل على وجود الارتباط الخطي. بما أن المصفوفة  $X'X$  في شكل مصفوفة الارتباط فإن المدي الممكن لقيم المحدد هو  $0 \leq |X'X| \leq 1$  ، عندما  $|X'X| = 1$  فإن المتغيرات المستقلة سوف تكون متعامدة بينما إذا كان  $|X'X| = 0$  فإنه يوجد علاقة خطية تامة بين المتغيرات المستقلة. الخطورة هو قرب  $|X'X|$  من الصفر. يعتبر هذا المقياس سهل في التطبيق ولكنه لا يمننا بأي معلومات على مصدر الارتباط الخطي. الآن سوف نقدم اختبارين الأول يعتمد على المحدد  $|X'X|$ ، أما الثاني فيعتمد على بعض المعايير الأحصائية.

(١) اختبار فرايبير - كلوبير - Farrar-Glauber :

ويعتمد هذا الاختبار على إحصاء مربع كاي  $(\chi^2)$  ويحسب من الصيغة التالية:

$$\chi_0^2 = \left[ n - 1 - \frac{1}{6}(2k + 5) \right] \ln |c^*|$$

حيث  $n$  حجم العينة و  $k$  عدد المتغيرات المستقلة و  $|c^*|$  اللوغاريتم الطبيعي لمحدد مصفوفة معاملات الارتباط التالية:

$$XX' = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1k} \\ r_{21} & 1 & r_{23} & \dots & r_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{k1} & r_{k2} & r_{k3} & \dots & 1 \end{bmatrix}$$

فرض العدم سوف يكون  $X_i$  متعامدة ضد الفرض البديل  $X_i$  غير متعامدة.  
وتقارن قيمة  $\chi^2_0$  المحسوبة مع قيمة  $\chi^2$  الجدولية في الملحق (٩) بدرجة حرية  $(k(k-1)/2)$ . إذا كانت القيمة المحسوبة اكبر من الجدولية نرفض فرض العدم ونقبل الفرض البديل.

مثال (٩-٤)

عينة عشوائية حجم (10) مشاهدات جمعت فيها البيانات عن كل من المتغير التابع (Y) وعلاقته بأربعة متغيرات مستقلة  $x_1, x_2, x_3, x_4$  وكما في جدول (٩-٦).

والمطلوب : اختبار وجود مشكلة الارتباط الخطي المتعدد.

جدول (٩-٦)

y	$x_1$	$x_2$	$x_3$	$x_4$
6.0	40.1	5.5	108	63
6.0	40.3	4.7	94	72
6.5	47.5	5.2	108	86
7.1	49.2	6.8	100	100
7.2	52.3	7.3	99	107
7.6	58.0	8.7	99	111
8.0	61.3	10.2	101	114
9.0	62.5	14.1	97	116
9.0	64.7	17.1	93	119
9.3	66.8	21.3	102	121



**الحل**

ولغرض إجراء اختبار العلاقة الخطية ، يجب حساب محدد مصفوفة الارتباطات البسيطة بين المتغيرات المستقلة.

$$XX' = \begin{bmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{21} & 1 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{bmatrix}$$

حيث:

$$XX' = \begin{bmatrix} 1 & 0.879 & -0.339 & 0.956 \\ 0.879 & 1 & -0.305 & 0.761 \\ -0.339 & -0.305 & 1 & -0.414 \\ 0.956 & 0.761 & -0.414 & 1 \end{bmatrix}$$

$$|c^*| = 0.0089 \quad \text{وبالتالي فإن:}$$

لذلك فإن قيمة مربع كاي المحسوبة تساوي:

$$\begin{aligned} \chi_0^2 &= - \left[ 10 - 1 - \frac{1}{6}(8 + 5) \right] \ln(0.0098) \\ &= (-6.8333) (-4.6253729) \\ &= 31.606699 \end{aligned}$$

وبمقارنة هذه القيمة مع قيمة مربع كاي  $(\chi^2)$  الجدولية لدرجة حرية مساوية (6) ومستوي معنوية  $\alpha = 0.05$  والمساوية إلى 12.592. لأن:

$$31.606699 > 12.592.$$

ومنه نرفض فرضية العدم  $(H_0)$  ونقبل الفرضية البديلة  $(H_1)$  أي وجود مشكلة الارتباط الخطي المتعدد بين المتغيرات للنموذج الخطي المدروس.

**(ب) طريقة فريش المعدلة:**

ويتلخص هذه الطريقة في الخطوات التالية:

١- الحصول على معادلات الانحدار البسيطة بين المتغير التابع وكل من المتغيرات المستقلة على حدة.

٢- اختبار النتائج المتحصل عليها في ضوء المعايير الإحصائية القبلية.

٣- اختيار المعادلة المتحصل عليها في ضوء المعايير الإحصائية القبلية.

٤- وبلي ذلك إضافة المتغيرات مع اختبار آثارها على المعالم وأخطائها المعيارية ومعامل الارتباط المتعدد، فإذا زادت قيمة معامل الارتباط نتيجة إضافة المتغير الجديد دون أن تتحول أي من المعالم إلى معلمة غير مقبولة على أساس الاعتبارات القبلية كان هذا التغير مفيداً، وأضيف إلى المعادلة كمتغير مستقل. إما إذا لم يطرأ تغيير على قيمة معامل الارتباط، ولم يؤثر المتغير المضاف على قيم المعالم حذف هذا المتغير من بين المتغيرات المستقلة. وإذا أثر المتغير الجديد على إشارات وقيم المعالم فصارت غير مقبولة على أساس الاعتبارات النظرية القبلية، دل ذلك على وجود الارتباط الخطي فالمتغير الجديد له أهميته، ولكن بسبب الارتباط بينه وبين المتغيرات المستقلة الأخرى فلا يمكن إظهار أثره باستخدام طريقة المربعات الصغرى العادية. كما لا يعني ذلك ضروري حذفه من النموذج حتى لا يصل بنا هذا الحذف إلى توصيف خاطئ. ولتصبح مثل هذا الوضع يمكننا اتباع إحدى طرق معالجة الارتباط الخطي التي سنشرحها فيما بعد.

للمثال (٩-٢) المطلوب إيجاد تقدير لنموذج الانحدار:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon.$$

وباستخدام طريقة المربعات الصغرى فإننا نحصل على النتائج التالية:

$$\hat{y} = -13.53 + 0.097x_1 - 0.199x_2 + 0.015x_3 + 0.34x_4,$$

$$R^2 = 0.998.$$

ولاختبار فرض العدم:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0,$$

نصيب قيمة F كالآتي:

$$F = \frac{MSR(\beta_1, \beta_2, \beta_3, \beta_4 | \beta_0)}{MSE(\beta_1, \beta_2, \beta_3, \beta_4 | \beta_0)} \\ = \frac{41.542}{0.06631} = 626.463.$$

ولما كانت قيمة F المحسوبة تزيد عن القيمة الجدولية عند درجات حرية (4,5) حيث  $F_{0.05}(4, 5) = 5.19$  وانلك نرفض فرض العدم وتقبل الفرض البديل بأن العلاقة بين الاتفاق على الملابس وبإلبي المتغيرات المستقلة علاقة معنوية مصفوفة الارتباط الخطي هي :

$$X'X = \begin{bmatrix} 1 & 0.98 & 0.988 & 0.988 \\ & 1 & 0.970 & 0.992 \\ & & 1 & 0.969 \\ & & & 1 \end{bmatrix}$$

وللبحث عن آثار العلاقة الخطية ، نحسب معادلات الانحدار البسيطة بين الاتفاق على الملابس وكل من المتغيرات المستقلة على حده. وفيما يلي نتائج هذه المعادلات والقيم بين الأقواس تمثل تقدير للأخطاء المعيارية للمقدرات.

$$\begin{aligned} \text{MSE} &= 0.09424 \quad R^2 = 0.995 \\ \hat{y} &= -1.246 + 0.118x_1 \\ &\quad (0.376) \quad (0.003) \end{aligned} \quad (1)$$

$$\begin{aligned} \text{MSE} &= 0.103 \quad R^2 = 0.996 \\ \hat{y} &= 1.405 + 0.126x_1 - 0.0361x_2 \\ &\quad (4.926) \quad (0.015) \quad (0.067) \end{aligned} \quad (2)$$

$$\begin{aligned} \text{MSE} &= 0.112 \quad R^2 = 0.996 \\ \hat{y} &= 0.94 + 0.139x_1 - 0.0345x_2 - 0.0379x_3 \\ &\quad (5.180) \quad (0.025) \quad (0.070) \quad (0.057) \end{aligned} \quad (3)$$

$$\begin{aligned} \text{MSE} &= 0.05629 \quad R^2 = 0.998 \\ \hat{y} &= -12.759 + 0.104x_1 - 0.188x_2 + 0.319x_4 \\ &\quad (6.516) \quad (0.014) \quad (0.076) \quad (0.122) \end{aligned} \quad (4)$$

$$\begin{aligned} \text{MSE} &= 0.06631 \quad R^2 = 0.998 \\ \hat{y} &= -13.53 + 0.0965x_1 - 0.199x_2 + 0.01508x_3 + 0.34x_4 \\ &\quad (7.513) \quad (0.026) \quad (0.09) \quad (0.0549) \quad (0.15) \end{aligned} \quad (5)$$

وتكون الخطوة الأولى في اختبار معادلة الانحدار الأولى حيث أن الدخل التصرفي يعتبر أكثر المتغيرات المستقلة أهمية خلال فترة الدراسة. يعطى جدول (٧-٩) نتائج الإضافة.

جدول (٧-٩)

$b_0$	$b_1$	$b_2$	$b_3$	$b_4$	$R^2$
-1.246 (0.376)	0.118 (0.003)	-	-	-	0.995
1.405 (4.926)	0.126 (0.015)	-0.036 (0.067)	-	-	0.996
0.94 (5.180)	0.139 (0.025)	-0.0345 (0.070)	-0.0379 (0.057)	-	0.996
-12.759 (6.516)	0.104 (0.014)	-0.188 (0.076)	-	0.319 (0.122)	0.998
-13.53 (7.513)	0.0965 (0.026)	-0.199 (0.09)	0.01508 (0.0549)	0.34 (0.15)	0.998

ويتضح من النتائج في جدول (٧-٩) أن الدخل له أهمية في شرح المتغيرات في الانفاق على الملابس ، وبإضافة  $x_2$  زادت قيمة  $R^2$  قليلا وكانت اشارات المعامل صحيحة والخطأ المعياري للمعلمة  $\beta_2$  يدل على عدم معنويتها الى جانب أن الارتباط القوي بين  $x_1$ ،  $x_2$  لم يؤثر على معنوية المعلمة  $\beta_1$  . أما إضافة  $x_3$  (الأصول السائلة) ، فقد أثر على تقديرات كل من  $\beta_3$ ،  $\beta_2$  فصارت غير مقبولة مما يدل على الارتباط القوي بين  $x_2$  ،  $x_3$  هو الذي أدى الى ذلك ولو أن  $b_1$  لم تتأثر بالرغم من الارتباط القوي بين  $x_1$ ،  $x_2$ ،  $x_3$  . ولذا كان من الأفضل حذف  $x_3$  . وبإضافة  $x_4$  تحسنت النتائج إذ زادت  $R^2$  قليلا وصارت جميع اشارات المعامل صحيحة ومعنوية احصائيا ، وبالرغم من الارتباط القوي بين المتغيرات المستقلة فإن الأخطاء المعيارية لمعاملها ليست كبيرة. وعندما حسب الانحدار للمتغيرات الأربعة دلت النتائج على أن الارتباط الخطي لم يؤثر على كل من  $b_1$ ،  $b_2$  بينما كانت  $b_3$  معنوية مما يؤكد ضرورة حذف المتغير  $x_3$  ، الأصول السائلة، وإذا يكون أفضل شكل للنموذج هو:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4 + \varepsilon.$$

### (٩-٦) معالجة الارتباط الخطي المتعدد

يتوقف أساليب معالجة الارتباط الخطي بين المتغيرات المستقلة ، إذا وجد في احدي نماذج الانحدار ، على درجة هذا الارتباط ومدى توفر البيانات وأهمية المتغيرات التي تسببت في هذا الارتباط واخيرا الغرض من إيجاد تقدير للنموذج.

ويرى البعض امكان قبول المشكلة ان كان تأثيرها بسيطا على تقديرات المعالم ، ويقترح البعض حذف المتغيرات غير الهامة من النموذج ان ظهر تأثيرها بسبب وجود الارتباط الخطي. اما إذا كان للارتباط الخطي أثره الواضح على تقديرات معالم المتغيرات الهامة فلا بد من اتباع احدي الطرق الآتية للتصحيح.

١- زيادة حجم العينة حيث يؤدي ذلك إلى خفض قيم الأخطاء المعيارية لمعالم النموذج.

٢- استخدام معلومات قبلية حول العلاقات بين المتغيرات المستقلة فمثلا نجد أن هناك علاقة بين المرتبة الوظيفية ومدة خبرة الموظف في العمل فبدلا من ادخال متغير المرتبة  $x_1$  والخبرة  $x_2$  متغيرين مستقلين ضمن متغيرات مستقلة أخرى للجبر الذي يتقاضاه الموظف  $Y$  فبممكن دمج هذين المتغيرين الى متغير واحد إذا أمكن الحصول على معلومات تقريبية تفيد بأن قيمة معلمة الخبرة مثلا تساوي حوالي ربع معلمة المرتبة وبالحصول على هذه المعلومة يتم بناء النموذج كالتالي:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$$

وبما أن  $\beta_2 = 0.25\beta_1$  فإن النموذج السابق يصبح :

$$\begin{aligned} Y &= \beta_0 + \beta_1 x_1 + 0.25\beta_1 x_2 + \varepsilon \\ &= \beta_0 + \beta_1 (x_1 + 0.25x_2) + \varepsilon. \end{aligned}$$

ومن عيوب هذا الحل صعوبة تحديد الأثر الفردي للمتغيرين على المتغير الفردي.

٣- تقليل عدد المتغيرات المستقلة ذات الارتباط المرتفع باستخدام تحليل المكونات الأساسية أو التحليل العنقلي. وتهدف هاتان الطريقتان الى تحويل المتغيرات المرتبطة الى عدد أقل تسمى بالعوامل في حالة التحليل العنقلي وبالمكونات الأساسية في حالة تحليل المكونات الأساسية

، بحيث يكون لكل عامل من هذه العوامل / المكونات دالة تربطه ببعض أو كل هذه المتغيرات. وعلى ذلك استخدام المتغيرات الجديدة غير المترابطة مع بعضها البعض كمتغيرات مستقلة جديدة للمتغير التابع. وللمزيد حول تحليل المكونات الأساسية والتحليل العاملي يمكن الرجوع إلى (Mardia et al 1979) regression . وسوف نناقش تحليل المكونات الأساسية في البند (٩-٧).

٤- استخدام انحدار الحافة ridge . ويتم في هذه الطريقة تعديل في طريقة المربعات الصغرى العادية بحيث تسمح بمقدرات منحازة لمعاملات الانحدار . وعندما ينحاز مقدار بمقدار بسيط فقط ويكون أكثر دقة بكثير من مقدار غير منحاز فقد يكون لمقدر الأفضل ، لأن احتمال قربه من القيمة الحقيقية للمعلمه سيكون أكبر بكثير من احتمال قربه المقدر الأخير المتحيز لقيمة المعلمة الحقيقية.

ليكن  $b^R$  هو متجه معاملات انحدار الحافة المعيارية من الدرجة  $k \times 1$  ونحصل عليه بحل المعادلة التالية:

$$(X'X + cI)b^R = X'y$$

أو

$$b^R = (X'X + cI)^{-1}X'y$$

حيث  $X'X$  مصفوفة ارتباط المتغيرات  $X$  و  $X'y$  قيمة معاملات الارتباط الخطي البسيط بين  $y$  وكل متغير من المتغيرات المستقلة و  $c$  هو ثابت التميز وتتراوح قيمته بين الصفر والواحد الصحيح و  $I$  مصفوفة الوحدة بدرجة  $k \times 1$ . ولإيجاد قيم معاملات الانحدار الاصلي نستخدم العلاقة التالية:

$$b_i = \frac{s_y}{s_i} b_i^R, \quad i = 1, 2, \dots, k$$

حيث:

$$s_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}},$$

$$s_i = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

حيث  $x_i, y_i$  القيم المشاهدة الاصليه.

$$b_0 = \bar{y} - b_1\bar{x}_1 - \dots - b_k\bar{x}_k.$$

بحسب عامل تضخم التباين لمعاملات الانحدار الحافة من المصفوفة التالية:

$$(11-9) \quad (X'X + cI)^{-1} X'y (X'y + cI)^{-1}.$$

ويعكس الثابت  $c$  مقدار الانحياز في المقدرات . وعندما  $c = 0$  تختزل (١١-٩) إلى معاملات انحدار المربعات الصغرى العادية ، وعندما يكون  $c > 0$  فإن معاملات انحدار الحافة تكون متحيزة ولكنها تميل إلى أن تكون أكثر استقراراً من مقدرات المربعات الصغرى. يعاب على طريقة انحدار الحافة صعوبة تحديد قيمة  $c$  المثلى. ولتحديد قيمة التحيز  $c$  التي تعطي أفضل نموذج يستخدم عادة الرسم البياني لقيم معاملات انحدار الحافة (المحور الرأسي) مع قيم مختلفة لثابت التميز ذات مسافات متساوية (المحور الأفقي). ويعرف الشكل الناتج بأثر الحافة . كما يؤخذ في الاعتبار قيمة عامل تضخم التباين للتأكد من حل مشكلة الارتباط.

وتشير الخبرة إلى إمكانية تذبذب معامل الانحدار المقدر  $b^R$  تذبذباً واسعاً عندما تتحرك  $c$  قليلاً عن الصفر بل يمكن أن تغير إشارتها، إلا أن هذه التذبذبات الواسعة تتوقف تدريجياً ويميل مقدار معامل الانحدار إلى التغير تغيراً بطيئاً فقط عندما يزداد  $c$  شيئاً فشيئاً. بينما تميل قيمة معامل تضخم التباين إلى الهبوط بسرعة عندما تتحرك  $c$  قليلاً عن الصفر وتميل قيمة معامل تضخم التباين بصورة تدريجية أيضاً إلى مجرد التغير باعتدال عند زيادة  $c$  شيئاً فشيئاً ولذلك تختار أصغر قيمة لـ  $c$  تبدو معها معاملات الانحدار وكأنها تستقر وللمره الأولى في أثر الحافة وتصبح معها قيم معامل تضخم التباين صغيرة صفراً كافياً. أي أن الاختيار هنا هو مسألة اجتهد.

#### مثال (٩-٥)

البيانات المعطاه في جدول (٨-٩) تمثل الواردات ( $y$ ) والنتائج القومي الاجاملى ( $x_1$ ) وكلها ببلاتين الدولارات ، والرقم القياسي العام لأسعار المستهلكين ( $x_2$ ) للولايات المتحدة الأمريكية من عام ١٩٦٤ إلى عام ١٩٧٩ والمطلوب تقدير نموذج الانحدار للواردات على النتائج القومي والرقم القياسي للأسعار والكشف عن وجود ارتباط خطي بين المتغيرين المستقلين واقتراح حلاً مناسباً لمشكلة الارتباط الخطي المتعدد إن وجدت.

جدول (٩-٨)

العام	الواردات (y)	النتائج القومي الإجمالي (x <sub>1</sub> )	الرقم القياسي للسعار (x <sub>2</sub> )
1964	28.4	635.7	92.9
1965	32.0	688.1	94.5
1966	37.7	753.0	97.2
1967	40.6	796.3	100.0
1968	47.7	868.5	104.2
1969	52.9	935.5	109.8
1970	58.5	982.4	116.3
1971	64.0	1063.4	121.3
1972	75.9	1171.1	125.3
1973	94.4	1306.6	133.1
1974	131.9	1412.9	147.7
1975	126.9	1528.8	161.2
1976	155.4	1702.2	170.5
1977	185.8	1899.5	181.5
1978	217.5	2127.6	195.4
1979	260.9	2368.5	217.4

الحل

نموذج الانحدار المقدر هو:

$$\hat{y} = -101.489 + 0.07853 x_1 + 0.759 x_2$$

(33.080)      (0.056)      (0.761)

(.009)      (0.184)      (0.337)

حيث الأرقام بين الأقواس أسفل معاملات الانحدار المقدره هي تقديرات للأخطاء المعيارية للمقدرات وقيم المعنوية (p - Value) المناظرة لكل معامل. توضح



النتائج معنوية الانحدار ككل ( $p\text{-Value} = 0.00$ ) وأن النموذج يفسر 98.7% ( $R^2 = 0.987$ ) من التغير في الواردات خلال الفترة من ١٩٦٤ إلى ١٩٧٩. يتضح وجود ارتباط خطي بين  $x_1$ ,  $x_2$  وذلك للأسباب الآتية:

- بالرغم من معنوية الانحدار ككل وكبر حجم معامل التحديد إلا أن  $b_1, b_2$  وهما معاملا الناتج القوي والرقم القياسي للأسعار ليست معنوية حيث بلغت قيمتا الاحتمال 0.34, 0.18 على التوالي. وذلك دليل قوي لوجود ارتباط خطي بين المتغيرين المستقلين.

- بما أن قيمة معامل الارتباط البسيط بين  $x_1$  و  $x_2$  هو  $r_{12} = 0.997$  والذي يقترب من الواحد الصحيح فهذا يعني وجود الارتباط الخطي بين  $x_1$ ,  $x_2$ .

- عامل تضخم التباين للمتغير  $x_1$  هو  $VIF_1 = 176.64$  وعامل تضخم التباين للمتغير  $x_2$  هو  $VIF_2 = 176.64$ .

وبما أن قيمة عامل تضخم التباين لمعاملى الانحدار متساويين وقيمتهما أكبر بكثير من 10 وهذا يوضح حجم مشكلة الارتباط الخطي الخطير الذي يعاني منها هذا النموذج.

لملاج مشكلة الارتباط الخطي لهذا النموذج سوف نستخدم انحدار الحافه. مصفوفة الارتباط الخطي البسيط بين المتغيرات المستقلة هي:

$$(X'X)^{-1} = \begin{pmatrix} 1 & 0.99717 \\ 0.99717 & 1 \end{pmatrix}$$

وقيمة معاملات الارتباط الخطي البسيط بين المتغير التابع مع المتغيرين المستقلين هو:

$$X'y = \begin{pmatrix} 0.993177 \\ 0.992699 \end{pmatrix}$$

مصفوفة الوحدة من الدرجة 2 x 2 هي:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

ويضرب ثابت التميز  $c = 0.5$  في مصفوفة الوحدة نحصل على:

$$cI = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$$

وعلى ذلك:

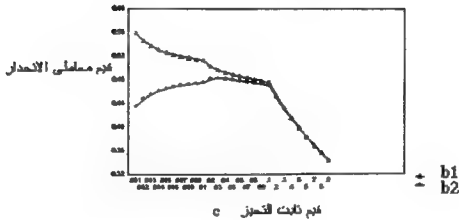
$$(X'X)^{-1} + cI = \begin{pmatrix} 1.5 & 0.99717 \\ 0.99717 & 1.5 \end{pmatrix}$$

$$\begin{aligned} ((X'X + cI)X'y)^{-1} &= \\ &\begin{pmatrix} 1.19459 & -0.79414 \\ -0.79414 & 1.19459 \end{pmatrix} \begin{pmatrix} 0.993177 \\ 0.992699 \end{pmatrix} \\ &= \begin{pmatrix} 0.398102 \\ 0.397151 \end{pmatrix}. \end{aligned}$$

وعلى ذلك فإن تقدير انحدار الحافة لمعالم النموذج هو:

$$b_1^R = 0.398102, \quad b_2^R = 0.397151.$$

يعطى جدول (٩-٩) قيم معاملات نموذج انحدار الواردات المقدره باستخدام طريقة انحدار الحافة لقيم مختلفة لثابت التميز وينضح من شكل (٩-١)



شكل (٩-١)

إن قيمتي معامل الانحدار تبدو مستقرة عند قيم  $c$  التي تتراوح بين 0.05 إلى 0.09. ويلاحظ ايضا ان عوامل تضخم التباين قد اخذت قيما اقل من الواحد الصحيح عند قيمة  $c = 0.05$  فأكبر. كما يجب ملاحظة انه من المرغوب فيه اختيار اصغر قيمة لـ  $c$  التي يحدث عندما الاستقرار. وفي هذا المثال تم اختيار القيمة  $c = 0.08$  كثابت تميز يعطى نموذجا لايحاشي من مشكلة الارتباط الخطي بين  $x_1, x_2$ .

جدول (۹-۱)

$R^2$	VIF	$b_2^R$	$b_1^R$	c
0.9874	96.328	0.434599	0.559251	0.001
0.9873	60.8875	0.447242	0.546111	0.002
0.9873	41.8826	0.45546	0.53739	0.003
0.9873	30.5915	0.461212	0.531148	0.004
0.9873	23.3397	0.465427	0.526437	0.005
0.9873	18.4078	0.468632	0.522737	0.006
0.9873	14.9027	0.471136	0.519739	0.007
0.9873	12.3222	0.473132	0.517249	0.008
0.9873	10.3679	0.474749	0.515138	0.009
0.9873	8.85205	0.476076	0.513318	0.010
0.9873	2.96368	0.481778	0.502711	0.020
0.9871	1.55767	0.482537	0.497095	0.030
0.9869	1.01311	0.481832	0.492991	0.040
0.9867	0.74601	0.480507	0.489554	0.050
0.9865	0.594955	0.478869	0.486477	0.060
0.9861	0.500867	0.477057	0.483619	0.070
0.9858	0.438007	0.47514	0.480911	0.080
0.9854	0.39369	0.473161	0.47831	0.090
0.9850	0.361079	0.471143	0.475791	0.100
0.97914	0.241302	0.45074	0.453096	0.200
0.97048	0.204689	0.431456	0.433034	0.300
0.959836	0.182510	0.41362	0.414807	0.400
0.947743	0.165742	0.397151	0.398101	0.500
0.934632	0.151942	0.38192	0.382713	0.600
0.920823	0.140137	0.367801	0.368481	0.700
0.906564	0.129828	0.354682	0.355278	0.800
0.892046	0.120709	0.342463	0.342992	0.900

(٧-٩) إحداد المكونات الرئيسية

**Principal Components Regression:**

يمكن الحصول على مقدرات متحيزة لمعاملات الإحداد وذلك باستخدام أسلوب يسمى إحداد المكونات الرئيسية. بفرض أن  $\Lambda$  مصفوفة قطرية من الدرجة  $k \times k$  حيث العناصر القطرية هي الجذور المميزة  $(\lambda_1, \lambda_2, \dots, \lambda_k)$  للمصفوفة  $X'X$  وعندما  $T$  تمثل مصفوفة المتجهات المميزة المرافقة للجذور المميزة فإن:

$$T'X'XT = \Lambda$$

وبوضع:

$$Z = XT,$$

$$\alpha = T'\beta$$

فإن نموذج الإحداد الخطي

$$y = X\beta + \epsilon$$

$$= (XT)(T'\beta) + \epsilon$$

يصبح:

$$= Z\alpha + \epsilon.$$

حيث:

$$Z'Z = T'X'XT = \Lambda, TT' = I.$$

الأعمدة للمصفوفة  $Z$  والتي تمثل فئة جديدة من متغيرات الإحداد المتعامدة بحيث أن:

$$Z = [Z_1, Z_2, \dots, Z_k]$$

تسمى المكونات الرئيسية.

مثال (٧-٩)

بفرض أن مصفوفة الارتباط هي:

$$X'X = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}$$

الجذور المميزة والمتجهات المميزة لها هي:

$$\lambda_1 = 1.4, \begin{bmatrix} t_{11} & 0.707 \\ t_{12} & 0.707 \end{bmatrix},$$

$$\lambda_2 = 0.6, \begin{bmatrix} t_{21} & 0.707 \\ t_{22} & -0.707 \end{bmatrix},$$

المصفوفة T سوف تكون:

$$T = \begin{bmatrix} 0.707 & 0.707 \\ 0.707 & -0.707 \end{bmatrix},$$

المصفوفة  $\Lambda$  سوف تكون:

$$\Lambda = \begin{bmatrix} 1.4 & 0 \\ 0 & 0.6 \end{bmatrix}$$

وعلى ذلك:

$$Z_1 = 0.707X_1 + 0.707X_2,$$

$$Z_2 = 0.707X_1 - 0.707X_2$$

تسمى المكونات الرئيسية.

مقدر المربعات الصغرى  $\alpha$  هو:

$$\hat{\alpha} = (Z'Z)^{-1} Z'y = \Lambda^{-1} Z'y$$

ومصفوفة التغاير لـ  $\hat{\alpha}$  هي:

$$\text{Cov}(\hat{\alpha}) = \sigma^2 (Z'Z)^{-1} = \sigma^2 \Lambda^{-1}$$

وعلى ذلك القيم الصغيرة من الجذور المميزة للمصفوفة  $XX'$  تعني أن تباین معاملات الانحدار المتعامدة والمقابلة لها سوف يكون كبير. غالبا يشار إلى الجذر المميز  $\lambda$  بأنه تباین المكون الرئيسي رقم  $\lambda$ . عندما تكون كل  $\lambda$  مساوية للواحد الصحيح، فهذا يعني أن متغيرات الانحدار الأصلية متعامدة، بينما إذا كان  $\lambda$  بالضبط يساوي الصفر، فهذا يعني أن علاقة خطية تامة بين المتغيرات الأصلية. في حالة وجود واحد أو أكثر من  $\lambda$  قريب من الصفر فهذا يعني وجود مشكلة الارتباط الخطي بين متغيرات الانحدار. مصفوفة التغاير لمقدرات معامل الانحدار القياسية  $B_i$  هي:

$$\text{Cov}(B) = \text{Cov}(T\hat{\alpha}) = T \Lambda^{-1} T' \sigma^2.$$

وهذا يعني أن  $\hat{B}_i$  تباين هو:

$$\text{Var} (B_i) = \sigma^2 \sum_{j=1}^k t_{ji}^2 / \lambda_j.$$

أي أن تباين  $\hat{B}_i$  يمثل علاقة خطية من معكوس الجذور المميزة. وهذا يوضح كيف أن واحد أو أكثر من الجذور المميزة الصغيرة القيمة تؤدي إلى عدم دقة تقدير المربعات الصغرى  $B_i$ .

علما مما سبق كيف أن الجذور المميزة ومتجهات الجذور المميزة المرافقة لها والخاصة بـ  $X'X$ . تمدنا بمعلومات خاصة عن طبيعة مشكلة الارتباط الخطي للمتغيرات المستقلة. وبما أن  $Z = XT$ ، فإن:

$$Z_i = \sum_{j=1}^k t_{ij} X_j \quad (12-9)$$

حيث  $X_j$  هو العمود رقم  $j$  من المصفوفة  $X$  و  $t_{ij}$  تمثل عناصر العمود  $i$  للمصفوفة  $T$  (أي المتجه المميز رقم  $i$  للمصفوفة  $X'X$ ). عندما يكون التباين للمكون الرئيسي رقم  $i$  ( $\lambda_i$ ) صغير، فهذا يعني أن  $Z_i$  تقترب من ثابت و (٩-١٢) تعني وجود علاقة خطية بين متغيرات الانحدار الأصلية تقترب من ثابت. وهذا هو تعريف مشكلة الارتباط الخطي بين متغيرات الانحدار. وعلى ذلك (٩-١٢) تفسر لنا لماذا عناصر المتجهات المميزة والمرتبطة بالجذور المميزة الصغيرة للمصفوفة  $X'X$  تعرف متغيرات الانحدار التي تسبب المشتركة في الارتباط الخطي.

يؤدي انحدار المكونات الرئيسية إلى التخلص من مشكلة الارتباط الخطي بين متغيرات الانحدار وذلك باستخدام فئة جزئية من المكونات الرئيسية في النموذج. للحصول على مقدار المكونات الرئيسية، نفترض أن متغيرات الانحدار مرتبة حسب الترتيب التنازلي للجذور المميزة،  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$ . ونفرض أن آخر  $s$  من تلك الجذور المميزة تقريبا مساوية للصفر. في انحدار المكونات الرئيسية فإن المكونات الرئيسية للجذور القريبة من الصفر تحذف من التحليل وتطبق طريقه المربعات الصغرى على المكونات الباقية والتي عددها  $k-s$ . وعلى ذلك فإن مقدار المكونات الرئيسية سيكون:

$$\hat{\alpha}_{kc} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \alpha_{k-s} \\ \hline 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \begin{array}{l} k-s \text{ مكون} \\ \\ s \text{ مكون} \end{array}$$

أو بدلالة المقدر القياسي:

$$b_{kc} = T \hat{\alpha}_{kc} = \sum_{j=1}^{k-s} \lambda_j^{-1} t_j' X_j' y t_j.$$

أن أسلوب اختبار المكونات الرئيسية التي عددها  $k-s$  يتم بحساب نسبته التباين الكلي الذي يعود إلى المكون الرئيسي رقم  $j$  كالتالي:

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_k}, j = 1, 2, \dots, k.$$

إذا كان من الممكن إرجاع الجزء الأكبر من التباين الكلي (80% أو 90% مثلاً) إلى المكونات الثلاثة الأولى (مثلاً) فإن تلك المكونات يمكن أن تحل محل متغيرات الانحدار الأصلية دون أن تفقد الكثير من المعلومات. للمثال (٩-٧) فإن نسبة التباين الكلي التي تعود إلى المكون الرئيسي الأول هو:

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{1.4}{1.4 + .6} = 0.7$$

أي أن المكون الرئيسي الأول يفسر 70% من التباين الكلي.

مثال (٩-٨)

يعطي جدول (٩-١٠) بيانات عن القيم المعيارية لمتغيرين مستقلين  $X_1$  و  $X_2$  ومتغير استجابة  $Y$ ، مع العلم أن البيانات الأصلية لمتغير الاستجابة: 20, 30, 33, 35, 63 : والبيانات الأصلية للمتغير المستقل الأول: 102, 104, 101, 93, 100 والثاني: 4, 5, 7, 1, 3.

جدول (٩-١٠)

y	x <sub>1</sub>	x <sub>2</sub>
-1.0092	0.4781	0
-0.3862	0.9562	0.4472
-0.1993	0.2390	1.3416
-0.0748	-1.6733	-1.3416
1.6695	0	-0.4472

والمطلوب: إيجاد المكونات الرئيسية ومعالجة الانحدار بدلالة القيم المعيارية ثم بدلالة القيم الأصلية.

الحل

مصفوفة معاملات الارتباط هي:

$$X'X = \begin{bmatrix} 1 & 0.7483 \\ 0.7483 & 1 \end{bmatrix}$$

الجنور المميزة للمصفوفة  $X'X$  هي:

$$\lambda_1 = 1.7483, \lambda_2 = 0.2517$$

قيم المتجه الأول المقابل للجنر المميز  $\lambda_1 = 1.7483$  سيكون:

$$\begin{bmatrix} 0.7072 \\ 0.7072 \end{bmatrix}$$

أما قيم المتجه الثاني المقابل للجنر المميز  $\lambda_2 = 0.2517$  فسيكون:

$$\begin{bmatrix} 0.7072 \\ -0.7072 \end{bmatrix}$$

المصفوفة T هي:

$$T = \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}$$



أي أن:

$$Z_1 = 0.7071X_1 + 0.7071X_2,$$

$$Z_2 = 0.7071X_1 - 0.7071X_2$$

المكون الرئيسي الأول يفسر وحده:

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{1.7483}{2} = 0.874$$

من التباين الكلي كما أن جذره المميز أكبر من واحد ( $\lambda_1 = 1.7483$ ) وعليه  
فيمكن إيجاد القيم المشاهدة للمكون الرئيسي الأول  $Z_1$  وذلك من البيانات في  
جدول (٩-١١).

جدول (٩-١١)

$x_1$	$x_2$	$z_1$	$y$
0.4781	0	0.3381	-1.0092
0.9562	0.4472	0.9923	-0.3862
0.2390	1.3416	1.1176	-0.1993
-1.6733	-1.3416	-2.1318	-0.0748
0	-0.4472	-0.3162	1.6695

وعلى ذلك للمشاهدة الأولى فإن:

$$z_{1j} = 0.7071(0.4781) + (0.7071)(0) = 0.3381.$$

نموذج الانحدار سوف يأخذ الشكل التالي:

$$y = \alpha_1 z_1 + \varepsilon$$

معادلة الانحدار المقدرة سوف تكون:

$$\hat{y} = -0.18814z$$

و معاملات الانحدار بدلالة القيم المعيارية سوف تكون:

$$b_{kc} = T\hat{\alpha}_{kc}$$

$$\hat{\alpha}'_{kc} = [-0.18814 \quad 0]$$

حيث:

وعلى ذلك:

- ٥٣٦ -

$$b'_1 = -0.13303, b'_2 = -0.13303$$

ومعاملات الانحدار الأصلية سوف تكون:

$$b_1 = -0.5104, b_2 = -0.95502, b_0 = -91.06.$$

## الفصل العاشر

### نماذج الانحدار الغير خطيه Nonlinear Regression Models

- (١-١٠) مقدمة
- (٢-١٠) نموذج الانحدار الغير خطي
- (٣-١٠) المربعات الصغرى الغير خطية
- (٤-١٠) التحويل إلى نموذج خطي
- (٥-١٠) تقدير المعالم في نظام غير خطي
- (٦-١٠) القيم المبدئية
- (٧-١٠) أمثلة للنماذج الغير خطية

### (١-١٠) مقدمة

يتناول هذا الفصل مقدمة مختصرة عن مشاكل التقدير للنماذج الغير خطية. لمزيد من المعلومات عن النماذج الغير خطية يمكن الرجوع إلى (Myers (1990) (Bates and Watts (1988) , (Draper and Smith (1981. في الفصول السابقة كان اهتمامنا بتوفيق نماذج الانحدار باستخدام طريقة المربعات الصغرى ، والتي تكون خطية في المعالم وعلى الشكل:

$$Y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \dots + \beta_k x_{kj} + \epsilon_j. \quad (1-10)$$

وبالرغم من أن النموذج (١-١٠) يمثل أنواع عديدة من العلاقات فإن هناك حالات يكون فيها هذا النموذج غير مناسب. على سبيل المثال ، إذا كانت هناك معلومات متوفرة عن شكل العلاقة بين المتغير التابع والمتغيرات المستقلة لا تمثل بالنموذج (١-١٠).

### (٢-١٠) نموذج الانحدار الغير خطي

#### Nonlinear Regression Model

يكتب نموذج الانحدار الغير خطي على الصورة التالية:

$$Y_j = f(x_j, \theta) + \epsilon_j, \quad j = 1, 2, \dots, n. \quad (2-10)$$

حيث حد الخطأ العشوائي له  $E(\epsilon_j) = 0$  ,  $Var(\epsilon_j) = \sigma^2$  . عادة يفترض أن  $\epsilon_j$  يتبع التوزيع الطبيعي. الدالة  $f$  هي دالة التوقع أو نموذج انحدار المجتمع حيث  $x_j$  متجه من متغيرات الانحدار و  $\theta$  متجه من الدرجة  $px_1$  من المعالم الغير معلومة. أيضا يلاحظ أن حد الخطأ تجميعي additive. يلاحظ أن هناك تشابه كبير بين النموذج (٢-١٠) والنموذج الخطي (١-١٠) فيما عدا أن  $E(Y_j)$  دالة غير خطية في المعالم.

في نماذج الانحدار الغير خطية فإن واحد على الأقل من مشتقات دالة التوقع بالنسبة للمعالم تعتمد على واحد على الأقل من المعالم. لتوضيح هذه النقطة و بغرض نموذج الانحدار الخطي:

$$Y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \dots + \beta_k x_{kj} + \epsilon_j.$$

بدالة توقع:

$$f(\underline{x}_j, \beta) = \beta_0 + \sum_{i=1}^k \beta_i x_{ij}.$$

الآن:

$$\frac{\partial f(\underline{x}_j, \beta)}{\partial \beta_i} = x_{ij}, \quad i = 0, 1, 2, \dots, k.$$

حيث  $x_{0j} = 1$ .

تذكر أنه في الحالة الخطية فإن المشتقات لا تكون دوال في المعالم .

الآن بفرض نموذج الانحدار الغير خطي:

$$\begin{aligned} Y_j &= f(x_j, \theta) + \epsilon_j \\ &= e^{-\theta x_j} + \epsilon_j \end{aligned}$$

وبما أن المشتقة لدالة التوقع ،  $e^{-\theta x_j}$  ، بالنسبة لـ  $\theta$  هي  $-x_j e^{-\theta x_j}$  ، أي دالة في  $\theta$  ، فإن النموذج غير خطي . سوف نستخدم الرمز  $\theta$  للمعالم في النماذج الغير خطية وذلك للتمييز بين الحالة الخطية والغير خطية.

### (٣-١٠) المربعات الصغرى الغير خطية

دالة المربعات الصغرى لنموذج انحدار غير خطي تكون على الشكل التالي:

$$S(\hat{\theta}) = \sum_{j=1}^n [y_j - f(x_j, \hat{\theta})]^2 \quad (٣-١٠)$$

حيث  $\hat{\theta}$  مقدر المربعات الصغرى للمعلمة  $\theta$  والذي يؤدي إلى تصغير  $S(\hat{\theta})$  .  
 بافتراض الاعتدال لحدود الخطأ في (٣-١٠) فإن  $\hat{\theta}$  أيضا يمثل مقدر الإمكان الأعظم للمعلمة  $\theta$  . في حالة نموذج الانحدار الخطي فإن هذا يؤدي إلى خواص جيدة للمقدرات ، على سبيل المثال الحصول على مقدر له أقل تباين . في الحالة الغير خطية فإننا لا نستطيع أن نضع أي جمل علمه عن خواص المقدرات فيما عدا إذا كانت حجوم العينات كبيره . وعلى ذلك اختبارات فروض وفترات ثقة تقريبية يمكن الحصول عليها.

للحصول على مقدر المربعات الصغرى  $\hat{\theta}$  فإننا نحتاج إلى الحصول على المشتقات الجزئية لـ (٣-١٠) بالنسبة لـ  $\hat{\theta}$  ، وهذا يؤدي إلى  $p$  من المعادلات

الطبيعية والتي لابد من حلها للحصول على  $\hat{\theta}$  . المعادلات الطبيعية تأخذ الشكل التالي:

$$\sum_{j=1}^n [y_j - f(x_j, \hat{\theta})] \left[ \frac{\partial f(x_j, \hat{\theta})}{\partial \theta_i} \right] = 0, i = 1, 2, \dots, p \quad (١-١٠)$$

حيث المقدار من القوسين هو المشتقة للدالة  $f(x_j, \hat{\theta})$  بالنسبة لـ  $\theta_i$  .

مثال (١-١٠)

بفرض أننا نرغب في الحصول على مقدر المربعات الصغرى  $\hat{\theta}$  للمعلمة  $\theta$  وذلك للنموذج  $Y_j = f(x_j, \theta) + \epsilon_j$  بمعلمه واحد حيث  $f(x_j, \theta) = e^{-\theta x_j}$  وذلك بالاعتماد على  $n$  من أزواج المشاهدات  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  .  
وبتطبيق المعادلة (١-١٠) نحصل على المعادلة الطبيعية التالية :

$$\sum_{j=1}^n [y_j - e^{-\theta x_j}] [-x_j e^{-\theta x_j}] = 0$$

أو:

$$\sum_{j=1}^n y_j x_j e^{-\theta x_j} - \sum_{j=1}^n x_j e^{-2\theta x_j} = 0.$$

يلاحظ أنه حتى في حالة نموذج انحدار غير خطي بسيط بمعلمة واحدة فإن الحصول على تقدير للمعلمة  $\hat{\theta}$  وذلك بحل معادلة طبيعية واحدة غير سهل. عند وجود أكثر من معلمة في نموذج الانحدار الغير خطي فإن حل المعادلات الطبيعية سوف يكون أصعب ، كما يتضح من المثال التالي :

مثال (٢-١٠)

بفرض نموذج الانحدار الغير خطي التالي:

$$Y_j = \exp(\theta_1 x_{1j}) + \exp(\theta_2 x_{2j}) + \epsilon_j .$$

وعلى ذلك لتصغير  $S(\hat{\theta})$  حيث :

$$S(\hat{\theta}) = \sum_{j=1}^n [y_j - \exp(\hat{\theta}_1 x_{1j}) - \exp(\hat{\theta}_2 x_{2j})]^2$$

وبتطبيق المعادلة (١-١٠) نحصل على المعادلتين التاليتين :

$$\sum_{j=1}^n [y_j - \exp(\hat{\theta}_1 x_{1j}) - \exp(\hat{\theta}_2 x_{2j})] x_{1j} \exp(\hat{\theta}_1 x_{1j}) = 0,$$

$$\sum_{j=1}^n [y_j - \exp(\hat{\theta}_1 x_{1j}) - \exp(\hat{\theta}_2 x_{2j})] x_{2j} \exp(\hat{\theta}_2 x_{2j}) = 0.$$

### (٤-١٠) التحويل إلى نموذج خطي

في بعض الاحيان يكون من المفيد استخدام تحويله تؤدي إلى تحويل النموذج الغير خطي إلى خطي. بفرض النموذج:

$$Y_j = \theta_1 e^{\theta_2 x_j} + \epsilon_j \quad (٥-١٠)$$

والذي يمكننا تحويل دالة التوقع له إلى خطيه وذلك بأخذ اللوغاريتم. وبالتالي يمكننا إعادة كتابة نموذج الاتحاد كالتالي:

$$Y_j = \ln \theta_1 + \theta_2 x_j + \epsilon_j$$

$$= \beta_0 + \beta_1 x_j + \epsilon_j \quad (٦-١٠)$$

حيث :

$$\theta_2 = \beta_1, \quad \beta_0 = \ln \theta_1.$$

وباستخدام نموذج الاتحاد الخطي البسيط يمكننا تقدير  $\beta_0, \beta_1$ . في بعض الأحيان فإن تقديرات المربعات الصغرى الخطية للمعالم في (٦-١٠) سوف لا تكافئ تقديرات المعالم الغير خطيه في النموذج الأصلي (٥-١٠). والسبب في ذلك أن طريقة المربعات الصغرى الأصلية تؤدي إلى تصغير مجموع مربعات البواقي على  $y$  بينما في النموذج المحول (٦-١٠) فإننا نصغر مجموع مربعات البواقي على  $\ln y$ .

ومما يجدر الإشارة إليه أن حد الخطأ في (٥-١٠) تجميعي additive ، وعلى ذلك أخذ اللوغاريتم على (٥-١٠) لن يؤدي إلى النموذج (٦-١٠). عندما يكون حد الخطأ مضروب في النموذج مثل:

$$Y_j = \theta_1 e^{\theta_2 x_j} \epsilon_j \quad (٧-١٠)$$

فإن أخذ لوغاريتم الطرفين يؤدي إلى :

$$\ln Y_j = \ln \theta_1 + \theta_2 x_j + \ln \epsilon_j$$

$$= \beta_0 + \beta_1 x_j + \epsilon_j^* \quad (٨-١٠)$$

وعندما  $\epsilon_j^*$  يتبع التوزيع الطبيعي فإن كل الخواص والاستدلالات لنموذج الانحدار الخطي البسيط سوف تطبق هنا. النموذج في (١٠-٧) يسمى نموذج قابل للتحويل إلى خطي.

مثال (١٠-٣)

بفرض النموذج الغير خطي التالي:

$$y_j = \frac{\theta_1}{\theta_2 + x_j} + \epsilon_j \quad (١٠-٩)$$

والمسمى معادلة Michaelis-Menten (1913) والتي استخدمت لوصف العلاقة بين متغيرين في مجال الكيمياء ، حيث استخدمت هذه المعادلة لسنوات عديدة في تقدير المعالم في kinetics .

يعطي جدول (١٠-١) بيانات لقيم  $x, y$  والمطلوب إيجاد تقدير لكل من  $\theta_1, \theta_2$ .

جدول (١٠-١)

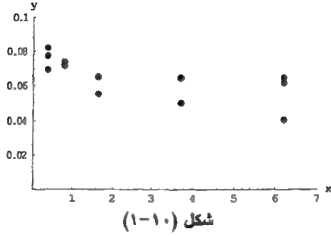
x	y
0.417	0.0773895
0.417	0.0688714
0.417	0.0819351
0.833	0.0737034
0.833	0.0738753
0.833	0.0712396
1.67	0.065042
1.67	0.0547667
3.75	0.0497128
3.75	0.0642727
6.25	0.0613005
6.25	0.0643576
6.25	0.0393892

المصدر:

Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1983.

شكل الانتشار للبيانات المعطاة في جدول (١٠-١) معطى في شكل (١٠-١).





يمكن تحويل دالة التوقع للنموذج (١٠-١) إلى خطية وذلك كالتالي:

$$\begin{aligned} \frac{1}{f(x_j, \theta)} &= \frac{\theta_2 + x_j}{\theta_1} \\ &= \frac{\theta_2}{\theta_1} + \frac{x_j}{\theta_1} \\ &= \beta_0 + \beta_1 x_j \end{aligned} \quad (١٠-١٠)$$

وعلى ذلك يمكن توفيق النموذج الخطي التالي :

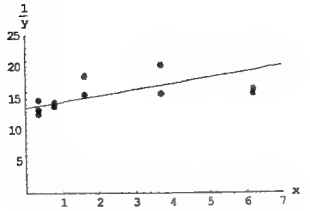
$$\begin{aligned} Y_j^* &= \beta_0 + \beta_1 x_j + \epsilon_j \\ (\text{حيث } Y_j^* &= 1/Y_j) \end{aligned}$$

معادله الانحدار المقدرة للنموذج (١٠-١) هي:

$$\hat{y}^* = 13.4306 + 0.981916x.$$

والموضحة بيانيا في شكل (١٠-٢) مع شكل الانتشار للبيانات المحولة  $y_j^*$ .

- ٥٤٤ -



شكل (١٠-٢)

وبما أن

$$\beta_0 = \frac{\theta_2}{\theta_1},$$

$$\beta_1 = \frac{1}{\theta_1}$$

فإن:

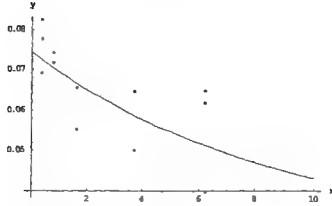
$$\hat{\theta}_1 = \frac{1}{b_1} = \frac{1}{0.981916} = 1.0184,$$

$$\hat{\theta}_2 = b_0 \hat{\theta}_1 = (13.4306)(1.0184) \\ = 13.677.$$

وعلى ذلك :

$$\hat{y} = \frac{1.0184}{13.677 + x}$$

والممثلة بيانيا في شكل (١٠-٣) مع الانتشار للبيانات الأصلية المعطاة في جدول (١٠-١).



شكل (١٠-٣)

### (١٠-٥) تقدير المعامل في نظام غير خطي

#### Parametric Estimation in a Nonlinear System.

الطريقة الأكثر انتشاراً في البرامج الجاهزة على الحاسب الآلي والمتخصصة في الانحدار الغير خطي هي طريقة التكرارات لـ جاكوبس - نيوتن Gauss-Newton . وفيها يتم الوصول إلى الخطية وذلك باستخدام مفكوك تيلور

للدالة  $f(x_j, \theta)$  عند النقطة  $\theta_0 = (\theta_{01}, \theta_{02}, \dots, \theta_{0p})$  حيث:

$$f(x_j, \theta) = f(x_j, \theta_0) + \sum_{i=1}^p \left[ \frac{\partial f(x_j, \theta)}{\partial \theta_i} \right]_{\theta=\theta_0} (\theta_i - \theta_{0i}) \quad (١٠-١١)$$

المعادلة (١٠-١١) يمكن وضعها على الصورة التقريبية التالية:

$$Y_j - f_j^0 = \sum_{i=1}^p \beta_i^0 z_{ij}^0 + \epsilon_j, \quad j = 1, 2, \dots, n. \quad (١٠-١٢)$$

حيث:

$$f_j^0 = f(x_j, \theta_0),$$

$$\beta_i^0 = \theta_i - \theta_{0i},$$

$$z_{ij}^0 = \left[ \frac{\partial f(x_j, \theta)}{\partial \theta_i} \right]_{\theta=\theta_0}$$

الآن حصلنا على نموذج انحدار خطي. عادة تسمى  $\theta_0$  القيم المبدئية starting values للمعامل. سوف نكتب (١٠-١٢) على الصيغة التالية:

$$Y_0 = Z_0 \beta_0 + \epsilon \quad (١٠-١٣)$$

حيث التقدير لـ  $\beta_0$  هو :

$$\underline{b}_0 = (Z_0' Z_0)^{-1} Z_0' \underline{y}_0 ,$$

$$Z_0 = \begin{bmatrix} z_{11}^0 & z_{21}^0 & \dots & z_{p1}^0 \\ z_{12}^0 & z_{22}^0 & \dots & z_{p2}^0 \\ \dots & \dots & \dots & \dots \\ z_{1j}^0 & z_{2j}^0 & \dots & z_{pj}^0 \\ \dots & \dots & \dots & \dots \\ z_{1n}^0 & z_{2n}^0 & \dots & z_{pn}^0 \end{bmatrix}$$

حيث  $Z_0$  مصفوفة من الدرجة  $n \times p$  و

$$\underline{b}_0 = \begin{bmatrix} b_1^0 \\ b_2^0 \\ \dots \\ b_p^0 \end{bmatrix} , \underline{y}_0 = \begin{bmatrix} y_1 - f_1^0 \\ y_2 - f_2^0 \\ \dots \\ y_j - f_j^0 \\ \dots \\ y_n - f_n^0 \end{bmatrix} \quad (١٠-١٤)$$

وبما أن  $\underline{b}_0 = \underline{\theta} - \underline{\theta}_0$  فإنه يمكن تعريف:

$$\underline{\hat{\theta}}_1 = \underline{b}_0 + \underline{\theta}_0 \quad (١٠-١٥)$$

كتقدير محسن لـ  $\underline{\theta}$  . الآن نضع التقدير المحسن  $\underline{\hat{\theta}}_1$  في (١٠-١١) ثم نحصل على فئة أخرى من التقديرات المحسنة ولتكن  $\underline{\hat{\theta}}_2$  وذلك بنفس القاعدة التي استخدمناها مع التقديرات  $\underline{\theta}_0$  ؛ حيث:

$$\underline{\hat{\theta}}_2 = \underline{\hat{\theta}}_1 + \underline{b}_1$$

عموماً عند التكرار رقم  $k$  فإن:

$$\underline{\hat{\theta}}_{k+1} = \underline{\hat{\theta}}_k + \underline{b}_k$$

$$= \underline{\hat{\theta}}_k + (Z_k' Z_k)^{-1} Z_k' (y - f_k) \quad (١٠-١٦)$$

حيث:

$$Z_k = \{z_{ij}^k\}, \quad y = [y_1, y_2, \dots, y_n]',$$

$$f_k = [f_1^k, f_2^k, \dots, f_n^k]',$$

$$\hat{\theta}_k = [\hat{\theta}_{k1}, \hat{\theta}_{k2}, \dots, \hat{\theta}_{kp}]'$$

وتستمر عملية التكرارات حتى التقارب convergence ، اي عندما:

$$|(\hat{\theta}_{k+1,i} - \hat{\theta}_{ki}) / \hat{\theta}_{ki}| < \delta, i=1,2,\dots,p.$$

حيث  $\delta$  عدد صغير (ليكن  $1.0 \times 10^{-6}$ ). في كل محاولة فإننا نحسب مجموع مربعات البوالي وذلك للتأكد من تحقق الاختزال في قيمته. عندما نحصل على التقدير النهائي للمعلمة  $\theta$  ولتكن  $\hat{\theta}$  فإننا نحسب مجموع مربعات البوالي كالتالي:

$$MSE = s^2 = \frac{\sum_{j=1}^n [y_j - f(x_j, \hat{\theta})]^2}{n - p}$$

كتقدير للمعلمة  $\sigma^2$ . التقدير لمصفوفة التغاير التقريبية لـ  $\hat{\theta}$  هي:

$$\Sigma = s^2 (Z'Z)^{-1} \quad (١٧-١٠)$$

حيث  $Z$  هي المصفوفة للمشتقات التفاضلية والمعرفة سابقا والمقدرة بتقدير المربعات الصغرى التي تم الحصول عليها في نهاية التكرارات التي عددها  $k+1$ . طريقة التكرارات لنيوتن والتي تم مناقشتها قد تتقارب ببطء في بعض الحالات وتتطلب عدد كبير من التكرارات. في حالات أخرى قد تتحرك في اتجاه خطأ مع زيادة في مجموع مربعات البوالي أوقد تفشل عملية التقارب. كثير من التعديلات على طريقة جاكوس - نيوتن افترضت وذلك لتحسين الطريقة. في واحد من تلك الطرق يعتبر  $\hat{b}_k$  هو المتجه القياسي في (١٠-١٦) والمحاولة رقم  $k$  عندما فقط  $S(\hat{\theta}_{k+1}) < S(\hat{\theta}_k)$  ، و نستمر إلى المحاولة التالية. بينما إذا كان  $S(\hat{\theta}_{k+1}) > S(\hat{\theta}_k)$  فإننا نستخدم  $\hat{b}_k / 2$  كمتجه من الزيادات. إذا حدث انه بعد عدة مرات من المحاولات الخاصة فإن الاختزال في  $S(\hat{\theta}_{k+1})$  لا نحصل عليه فإن الطريقة توقف.

وهناك تحسين آخر لطريقة جاكوس - نيوتن والمقدمة من قبل Marquardt (1963) والتي تعتمد على حساب المتجه  $\hat{b}_k$  عند التكرار رقم  $k$  من المصفوفة التالية:

$$(Z_k' Z_k + \lambda I_p) \hat{b}_k = Z_k' (y - f_k)$$

حيث  $\lambda > 0$  . ومما يجدر الاشارة اليه ان هذا التحسين يشابه طريقة انحدار الجسر التي تناولناها في الفصل التاسع . وقد استخدم (1963) Marquardt بحث في إيجاد قيمة  $\lambda$  والتي تؤدي إلى اختزال مجموع مربعات البواقي عند كل مرحلة . كثير من برامج الحاسب الآلي تختار  $\lambda$  بطرق مختلفة .

مثال (١٠-٤)

استخدم (1988) Bates and Watts طريقة جاوس - ماركوف وذلك لتوفير النموذج الغير خطي التالي:

$$y_j = \frac{\theta_1 x_j}{x_j + \theta_2} + \epsilon_j$$

وذلك باستخدام القيمتين  $\theta_{01} = 205$  ,  $\theta_{02} = 0.08$  كقيم مبدئية . سوف نوضح فيما بعد كيف يمكن الحصول على تلك القيم المبدئية . عند هذه النقطة فإن  $S(\hat{\theta}_0) = 3155$  . البيانات اللازمة للحصول على تقديرات للمعالم  $\theta_1$  ,  $\theta_2$  معطاة في جدول (١٠-٢) .

جدول (١٠-٢)

j	$x_j$	$y_j$	$f_j^0$	$y_j - f_j^0$	$z_{1j}^0$	$z_{2j}^0$
1	0.02	76	41.00	35.00	0.2000	-410.00
2	0.02	47	41.00	6.00	0.2000	-410.00
3	0.06	97	87.86	9.14	0.4286	-627.55
4	0.06	107	87.86	19.14	0.4286	-627.55
5	0.11	123	118.68	4.32	0.5789	-624.65
6	0.11	139	118.68	20.32	0.5789	-624.65
7	0.22	159	150.33	8.67	0.7333	-501.11
8	0.22	152	150.33	1.67	0.7333	-501.11
9	0.56	191	179.38	11.62	0.8750	-280.27
10	0.56	201	179.38	21.62	0.8750	-280.27
11	1.10	207	191.10	15.90	0.9322	-161.95
12	1.10	200	191.10	8.90	0.9322	-161.95

لتوضيح كيف يمكن حساب المشتقات سوف نتبع التالي:

$$\frac{\partial f(x_j, \theta_1, \theta_2)}{\partial \theta_1} = \frac{x_j}{\theta_2 + x_j},$$

$$\frac{\partial f(x_j, \theta_1, \theta_2)}{\partial \theta_2} = \frac{-\theta_1 x_j}{(\theta_2 + x_j)^2}$$

وبما أن أول مشاهدته على  $x$  هي  $x_j = 0.02$  فإن :

$$z_{11}^0 = \frac{x_1}{\theta_2 + x_1} \Big|_{\theta_2 = 0.08} = \frac{0.02}{0.08 + 0.02} = 0.2000,$$

$$z_{21}^0 = \frac{-\theta_1 x_1}{(\theta_2 + x_1)^2} \Big|_{\theta_1 = 205, \theta_2 = 0.08} = \frac{(-205)(0.02)}{(0.08 + 0.02)^2} = -410.00.$$

الآن المشتقات  $z_{jj}^0$  المعطاة في جدول (١٠-٢) تكون المصفوفة  $Z_0$  والمتجه من الزيادات يقدر من المعادلة (١٠-١٤) كالآتي:

$$\underline{b}_0 = \begin{bmatrix} 8.03 \\ -0.017 \end{bmatrix}.$$

التقدير المحسن  $\hat{\theta}_1$  يقدر من (١٠-١٥) كالآتي:

$$\begin{aligned} \hat{\theta}_1 &= \underline{b}_0 + \underline{\theta}_0 \\ &= \begin{bmatrix} 8.03 \\ -0.017 \end{bmatrix} + \begin{bmatrix} 205.00 \\ 0.08 \end{bmatrix} = \begin{bmatrix} 213.03 \\ 0.063 \end{bmatrix}. \end{aligned}$$

مجموع مربعات البواقي عند هذه النقطة هي  $S(\hat{\theta}_1) = 1206$  والذي يعتبر اصغر من  $S(\hat{\theta}_0)$ . وباستمرار التكرارات فإن التقدير لـ  $\hat{\theta}$  سوف يكون:

$\hat{\theta} = [212.7, 0.641]'$  حيث  $S(\hat{\theta}_1) = 1195$  وذلك بدرجة حرية  $n - p = 10$  وعلى ذلك  $s^2$  سوف يكون:

$$s^2 = 119.5$$

مصفوفة التغاير التقريبية للمتجه  $\hat{\theta}$  سوف تكون:

$$\Sigma = s^2(Z'Z)^{-1}$$

$$= 119.5 \begin{bmatrix} 0.4037 & 36.82 \times 10^{-5} \\ 36.82 \times 10^{-5} & 57.36 \times 10^{-8} \end{bmatrix}.$$

وعلى ذلك خطأ معياري مقرب لمقدرات المعالم سوف يكون:

$$\sqrt{\text{Var}(\hat{\theta}_1)} = \sqrt{119.5(0.4037)} = 6.95,$$

$$\sqrt{\text{Var}(\hat{\theta}_2)} = \sqrt{119.5(57.36) \times 10^{-8}}$$

$$= 8.28 \times 10^{-3},$$

معامل الارتباط بين  $\hat{\theta}_2, \hat{\theta}_1$  هو:

$$\frac{36.82 \times 10^{-5}}{\sqrt{0.4037(57.36 \times 10^{-8})}} = 0.77.$$

95% فترات ثقة تقريبية لكل من  $\hat{\theta}_1, \hat{\theta}_2$  سوف تكون:

$$\hat{\theta}_1 \pm t_{0.025}(10)(6.95),$$

$$\hat{\theta}_2 \pm t_{0.025}(10)(8.28 \times 10^{-3})$$

أو

$$212.7 \pm 15.5$$

لـ  $\theta_1$  و

$$0.0641 \pm 0.0185$$

لـ  $\theta_2$  .

مثال (١٠-٥)

في تجربه لقياس درجة حرارة كوب من الشاي (Y) عند أزمنة مختلفة (x) تم الحصول على البيانات المعطاة في جدول (١٠-٣). والمطلوب توفير النموذج:



$$Y_j = \theta_1 + \theta_2 e^{\theta_3 x_j} + \epsilon_j$$

جدول (۱۰-۳)

x	y	x	y	x	y	x	y
0	70.86	1	68.71	2	66.67	3	64.73
4	63.25	5	61.57	6	60.25	7	58.74
8	57.6	9	56.17	10	54.94	11	53.82
12	52.64	13	51.7	14	50.64	15	49.81
16	48.85	17	48.04	18	47.24	19	46.45
20	45.8	21	45.03	22	44.27	23	43.64
24	43.01	25	42.27	26	41.78	27	41.05
28	40.57	29	39.97	30	39.37	31	38.9
32	38.31	33	37.84	34	37.37	35	36.71
36	36.33	37	35.79	38	35.41	39	34.98
40	34.53	41	34.18	42	33.81	43	33.39
44	33.05	45	32.72	46	32.38	47	32.04
48	31.82	49	31.48	50	31.15	51	30.92
52	30.59	53	30.63	54	30.03	55	29.81
56	29.59	57	29.36	58	29.14	59	28.81
60	28.59	61	28.09	62	28.25	63	28.03
64	27.81	65	27.7	66	27.48	67	27.37
68	27.15	69	27.04	70	26.82	71	26.7
72	26.69	73	26.37	74	26.26	75	26.15
76	26.04	77	25.82	78	25.71	79	25.71
80	25.49	81	25.38	82	25.27	83	25.16
84	24.59	85	24.95	86	24.94	87	24.73
88	24.62	89	24.51	90	24.39	91	24.28
92	24.17	93	24.17	94	24.06	95	23.95
96	23.84	97	23.73	98	23.73		

الحل

تم استخدام البرنامج الجاهز والخاص بتوفيق النماذج الغير خطيه والمحمل على برنامج Mathematica الخاص بالبرامج الإحصائية.

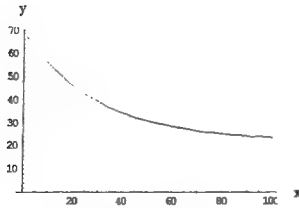
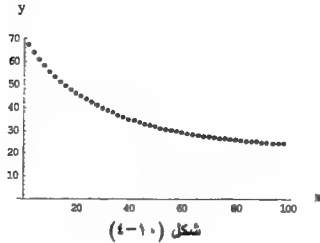
شكل الانتشار للبيانات المعطاه في جدول (٣-١٠) معطاه في شكل

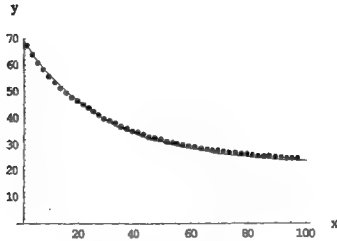
(٤-١٠). معادلة الانحدار المقدره هي:

$$\hat{y} = 21.978 + 47.0884 e^{0.033242x}$$

والممثلة بيانيا في شكل (٥-١٠). شكل الانتشار للبيانات المعطاه في جدول

(٣-١٠) مع معادلة الانحدار المقدره معطى في شكل (٦-١٠).





شكل (١٠-٦)

#### (١٠-٦) القيم المبدئية

يتطلب توفيق نموذج الانحدار الغير خطي معرفة قيم مبدئية  $\theta_0$  للمعالم النموذج. القيم الجيدة ، أي قيم  $\theta_0$  والتي تقترب من قيم المعالم الحقيقية سوف تؤدي إلى تصغير صعوبات التقارب. دائما الاختيار الجيد للقيم المبدئية يكون مفيد. في نماذج الانحدار الغير خطية فإن المعالم غالبا يكون لها معطي فيزيائي وهذا يساعد في اختيار القيم المبدئية للمعالم.

هناك تحويلات بيانية يمكن استخدامها والمقدمة من قبل Bates and Watts (1988) . في بعض الأحيان فإنه يمكن تحويل دالة التوقع للحصول على القيم المبدئية. على سبيل المثال في النموذج الغير خطي في مثال (١٠-٣) وتحويل النموذج في (١٠-٩) إلى نموذج خطي وذلك بأخذ معكوس دالة التوقع واستخدام طريقة المربعات الصغرى لتقدير المعالم حصلنا على قيم تقديرية  $\theta_1, \theta_2$  حيث  $\theta_1 = 1.0184, \theta_2 = 13.677$  ، والتي يمكن استخدامها في طريقة جلوس ... نيوتن كقيم مبدئية.

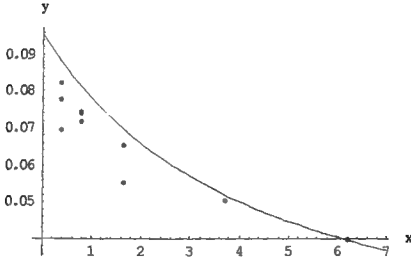
في المثال التالي سوف نوضح كيف أن معرفة القيم المبدئية للمنتج  $\theta$  سوف تساعدنا في الوصول إلى القيمة النهائية لتقدير  $\theta$  كما أن مجموع مربعات البواقي في هذه الحالة سوف يكون أقل من مجموع مربعات البواقي وذلك في حالة عدم معرفة القيم المبدئية للمنتج  $\theta$ .

مثال (١٠-٦)

للمثال (١٠-٣) تم استخدام برنامج جاهز خاص بالانحدار يتبع برنامج Mathematica . باستخدام  $\theta_{01} = 1$  ,  $\theta_{02} = 1$  كقيم مبدئية فإن معادلة الانحدار المقدرة سوف تكون:

$$\hat{y} = \frac{0.421171}{4.40374 + x}$$

والممثلة بيانيا مع شكل الانتشار في شكل (١٠-٧).

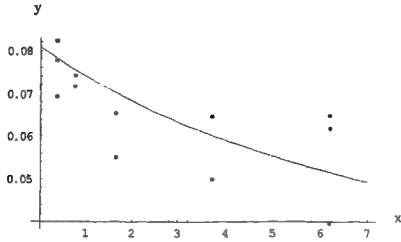


شكل (١٠-٧)

مجموع مربعات البواقي كانت 0.00212771. الآن بفرض أن هناك معلومات مبدئية عن  $\theta_1, \theta_2$  حيث  $\theta_{01} = 0.5, \theta_{02} = 17$  فإن معادلة الانحدار المقدرة هي:

$$\hat{y} = \frac{0.875918}{10.838 + x}$$

والممثلة بيانيا مع شكل الانتشار في شكل (١٠-٨).



شكل (١٠-٨)

مجموع مربعات البواقي كانت 0.000910677 والتي أقل من مجموع مربعات البواقي في حالة عدم معرفة القيم المبدئية. وهذا يعني أن الحل باستخدام القيم المبدئية  $\hat{\theta}_{01} = 0.5$  ,  $\hat{\theta}_{02} = 17$  كانت أكثر دقة . عدد التكرارات للوصول إلى الحل النهائي معطاة في جدول (١٠-٤).

جدول (١٠-٤)

i	$\hat{\theta}_{i1}$	$\hat{\theta}_{i2}$
0	0.5	17
1	0.997294	7.1388
2	0.924735	8.25897
.	0.9117	9.47141
.	0.90256	10.3259
.	0.885534	10.6872
.	0.877484	10.8141
.	0.876086	10.8354
.	0.875935	10.8377
.	0.87592	10.8379
.	0.875918	10.838

وبما أن  $s^2 = 0.00008279$  فإن مصفوفة التغاير التقريبية للمتجه  $\hat{\theta}$  سوف تكون:

$$\Sigma = s^2(Z'Z)^{-1}$$

$$= \begin{bmatrix} 0.0632335 & 0.886452 \\ 0.886452 & 12.6386 \end{bmatrix}.$$

وعلى ذلك خطأ معياري مقرب لمقدرات المعالم سوف يكون:

$$\sqrt{\text{Var}(\hat{\theta}_1)} = \sqrt{0.0632335} = 0.251463,$$

$$\sqrt{\text{Var}(\hat{\theta}_2)} = \sqrt{12.6386} = 3.55508.$$

مصفوفة معاملات الارتباط هي:

$$\begin{bmatrix} 1 & 0.991589 \\ 0.991589 & 1 \end{bmatrix}.$$

جدول تحليل التباين معطى في جدول (١٠-٥).

جدول (١٠-٥)

S.O.V	df	SS	MS	F
الانحدار	2	0.0557833	0.0278916	336.9
الخطأ	11	0.0009107	0.00008279	
الكلية	13	0.056694		

95% فترة ثقة لـ  $\theta_1, \theta_2$  هي:

(3.01329 , 18.6628) , (0.322452 , 1.42938) على التوالي.

مثال (١٠-٧)

النموذج التالي يصف الكثافة السكانية ( $Y_j$ ) في منطقة ريفية حيث:

$$Y_j = \theta_1 + \theta_2 d_j^{-\theta_3} + \epsilon_j$$

حيث  $d_j$  هي المسافة من مركز المدينة و  $\theta_3 > 0$  . عندما  $d_j$  كبيرة فإن

$\theta_1 \approx y_j$  . يمكن كتابة النموذج على الصورة التالية:

$$\log[y_j - \theta_1] = \log[\theta_2] - \theta_3 \log[d_j].$$

وعلى ذلك ، عندما  $\theta_1$  معلومة فإنه يمكن رسم  $\log[y_j - \theta_1]$  مقابل  $\log[d_j]$ .  
الجزء المقطوع سوف يكون  $\log[\theta_2]$  والميل هو  $\theta_3$ . تلك التقديرات سوف تكون قيمه مبدئية جيدة لتقدير المعالم.

#### (١٠-٧) أمثله للنماذج الغير خطيه

تستخدم النماذج الغير خطيه في مجالات كثيره وخصوصا في الكيمياء والفيزياء والعلوم الحيويه. ومن امثله النماذج الغير خطيه نماذج النمو والتي تستخدم كثيرا في مجال العلوم الحيويه حيث النباتات او الكائنات الدقيقه تنمو مع الزمن. المتغير المستقل هو الزمن. أيضا يستخدم نموذج النمو في كثير من مجالات الهندسة.

واحد من صيغ دالة النمو هو نموذج النمو اللوجستي :

$$Y_j = \frac{\theta_1}{1 + \theta_2 \exp(-\theta_3 x)} + \epsilon_j$$

عندما  $x=0$  فإن  $Y = \theta_1 / (1 + \theta_2)$  يمثل مستوى  $Y$  عند الزمن صفر. المعلمة  $\theta_1$  هي نهاية النمو عندما  $x \rightarrow \infty$ . القيم  $\theta_3, \theta_2$  لابد أن يكونان موجبين.  
هناك نموذج آخر للنمو يسمى Gompertz model والذي يأخذ الصيغة التالية:

$$Y_j = \theta_1 \exp(-\theta_2 e^{\theta_3 x}) + \epsilon_j.$$

عندما  $x=0$  فإن  $Y = \theta_1 e^{-\theta_2}$  ،  $\theta_1, Y$  نهاية النمو عندما  $x \rightarrow \infty$ .  
النموذج الاخير للنمو هو نموذج وايبل والذي يأخذ الصورة التالية:

$$Y_j = \theta_1 - \theta_2 \exp(-\theta_3 x^{\theta_4}) + \epsilon_j.$$

عندما  $x=0$  فإن  $Y = \theta_1 - \theta_2$  ، ونهاية النمو هو  $\theta_1$  عندما  $x \rightarrow \infty$ .

المراجع



المراجع :

**REFERENCES:**

**أولاً: المراجع العربية:**

٠١. أموري هادي كاظم ومحمد مناحد عيفان الدلمي ، (١٩٨٨م) ، " مقدمة في تحليل الانحدار الخطي " - وزارة التعليم العالي والبحث العلمي - جامعة بغداد.
٠٢. أنيس إسماعيل كنجو وآخرون ، (٢٠٠٠م) ، " نماذج إحصائية خطية تطبيقية - انحدار ، تحليل تباين وتصاميم تجريبية - الجزء الأول (الانحدار) " - ترجمة لكتاب جون نتر وآخرون - النشر العلمي والمطابع - جامعة الملك سعود.
٠٣. بدرية شوقي عبد الوهاب ومحمد كامل الشربيني ، (١٩٨٤م) ، " المبادئ الأولية في الإحصاء " - ترجمة لكتاب بول ج. هويل - الطبعة الرابعة - دار جون وايلي وأبنائه.
٠٤. ثروت محمد عبد المنعم ، (٢٠٠٤م) ، " مدخل حديث للإحصاء والاحتمالات " - الطبعة الثانية - مكتبة العبيكان.
٠٥. ثروت محمد عبد المنعم ، (٢٠٠٤م) ، " تصميم وتحليل التجارب " - مكتبة الانجلو المصرية.
٠٦. خاشع محمود الراوي ، (١٩٨٧م) ، " المدخل إلى تحليل الانحدار " ، وزارة التعليم العالي والبحث العلمي - جامعة الموصل - العراق.
٠٧. ربيع زكي عامر ، (١٩٨٩م) ، " تحليل الانحدار " - أساليبه وتطبيقاته العملية باستخدام البرامج الجاهزة SPSS/PC+ - معهد الدراسات والبحوث الإحصائية - جامعة القاهرة.
٠٨. سعد الدين محمد الشيال ، (١٩٧٦م) ، " مقدمة في الاقتصاد القياسي " - القاهرة - معهد الدراسات والبحوث الإحصائية.
٠٩. سعية حافظ منتصر ، (١٩٨٢م) ، " الإحصاء والاقتصاد القياسي " - ترجمة لكتاب دومنيك سالفاتور - دار ماكجروهيل للنشر - القاهرة - جمهورية مصر العربية.
١٠. عبد الحميد العباسي ، " التحليل والإحصاء باستخدام برنامج SPSS " - معهد الدراسات والبحوث الإحصائية - جامعة القاهرة.
١١. عبد الحميد عبد المجيد البلداوي ، (١٩٩٧م) ، " الإحصاء للعلوم الانسانية والتطبيقية " - دار القروق للنشر والتوزيع - عمان - الأردن.

١٢. عبد المرحضى حامد عزام ، (١٩٩٨م) . "التحليل الإحصائي للمتغيرات المتعددة من الوجه التطبيقية" - ترجمة لكتاب ريتشارد جونسون ودي وشرن - دار المريخ - الرياض - المملكة العربية السعودية.
١٣. محمد عبد الرحمن إسماعيل ، (٢٠٠١م) ، " تحليل الانحدار الخطي " - معهد الإدارة العامة - المملكة العربية السعودية - مركز البحوث.

### ثانياً: المراجع الأخرى:

1. Abell, M.L. et al,(1999), Statistics with Mathematical, Academic Press , New York .
2. Aitkin, M.A.,(1974), Simultaneous Inference and the Choice of Variable Subsets , Technometrics , 16,221-227.
3. Andrews, D.F.,(1974), A Robust Method for Multiple Linear Regression , Technometrics , 16, 523-531.
4. Andrews, D.F.,(1979), The Robustness of Residual Displays, In R.L. Launer and G.N. Wilkinson (Eds.), Robustness in Statistics, Academic Press, New York, pp. 19-32.
5. Anscombe, F.J.,(1960), Rejection of Outliers, Technometrics,2,123-167.
6. Anscombe, F.J. and Tukey, J.W.,(1963), The Examination and Analysis of Residuals , Technometrics ,5,141-160.
7. Bates, D.M. and Watts, D.G.,(1988), Nonlinear Regression Analysis and it's Applications , Wiley , New York.
8. Belsley, D.A. et al, (1980), Regression Diagnostics : Identifying Influential Data and Sources of Collinearity , Wiley , New York.
9. Box, G.E.P. and Tidwell, P.W., (1962) , Transformation of the Independent Variables , Technometrics, 4, 531-550.
10. Box, G.E.P. and Cox, D.R. ,(1964), An Analysis of Transformations , J.R. Statist. Soc. Ser.B,26,211-243.

11. Box, G. E. P. and . Wetz, J.M., (1973), Criterion of Judging the Adequacy of Estimation by an Approximating Response Polynomial , Technical Report No. 9, Department of Statistics, University of Wisconsin, Madison.
12. Carroll, R.J. and Ruppert, D. ,(1985), Transformation in Regression: A Robust Analysis, Technometrics, 27, 1-12.
13. Chatterjee, S. and Hadi, A.S.,(1988), Sensitivity Analysis in Linear Regression, New York , John Wiley.
14. Cook, R.D.,(1977), Detection of Influential Observations in Linear Regression , Technometrics , 19,15-18.
15. Cook, R.D.,(1979), Influential Observation in Linear Regression . J. Am. Statist. Assoc. ,74, 169-174.
16. Cook, R.D. and Weisberg, S.,(1980), Characterizations of an Empirical Influence Function for Detecting Influential Cases in Regression. Technometrics , 22, 495-508.
17. Daniel, C. and Wood, F. S., (1980), Fitting Equations to Data, 2<sup>nd</sup> Edition , Wiley, New York.
18. Devore, J.L.,(1995), Probability and Statistics for Engineering and the Sciences, 4<sup>th</sup> Edition , Duxbury Press- An International Thomson Publishing Company-London.
19. Draper, N.R. and Smith, H.,(1981), Applied Regression Analysis, 2<sup>nd</sup> Edition , Wiley , New York.
20. Efronson, M.A.,(1960), Multiple Regression Analysis , in A . Ralston and H.S. Wilf (Eds.) Mathematical Methods for Digital Computers, Wiley, New York.
21. Ellenberg, J.H.,(1976), Testing for a single Outlier From a General Regression , Biometrics, 32, 637-645.
22. Ezekiel, M.,(1930), Methods of Correlation Analysis , Wiley, New York.
23. Ezekiel, M. and Fox, K.A.,(1959), Methods of Correlation and Regression Analysis , Wiley , New York.
24. Fogiel, M. ,(1996), Problem Solvers, Statistics, Research and Education Association , New York.

25. Gnanadesikan, R. ,(1977), Method for Statistical Analysis of Multivariate Data, Wiley, New York.
26. Goldfeld, S.M. and Quandt , R.E.,(1965), Some tests for homoscedasticity, J. Am. Statist. Assoc. , 60 , 539-547.
27. Hahn, G.J., (1973), The Coefficient of Determination Exposed, Chem. Technol., 3, 609-614.
28. Land, C.E., (1974), Confidence Interval Estimation for Means After Data Transformation to Normality, J. Am. Statist. Assoc., 69, 795-802 (Correction, *ibid.* , 71, 255).
29. Larsen, W.A. and McCleary, S.J.,(1972),The Use of Partial Residual Plot In Regression Analysis, Technometrics , 14, 781-790.
30. Mallows, C.L.,(1973), Some Comments on Cp, Technometric,15,661-675.
31. Mardia, K.V., et al,(1979), Multivariate Analysis , Academic Press , London.
32. Marquardt, D.W., (1963), An Algorithm for Least Squares Estimation of Nonlinear Parameters , J. Soc. Ind. Appl. Math. ,2,431-441.
33. Marquardt, D.W.,(1970), Generalized Inverses Ridge Regression , Biased Linear Estimation , and Nonlinear Estimation , Technometrics , 12, 591-612.
34. Mason, R.L. et al ,(1975), Regression Analysis and Problems of Multicollinearity, Commun , Statist. ,4(3),277-292.
35. Montgomery, D.C. and Peck, E.A.,(1992), Introduction to Linear Regression Analysis, 2<sup>nd</sup> Edition , John Wiley, New York.
36. Mosteller, F. and Tukey , J.W.,(1977), Data Analysis and Regression : A Second Course in Statistics , Addison - Wesley , Reading , Mass.
37. Myers, R.H.,(1971), Response Surface Methodology , Aliyn and Bacon , Boston.

38. Myers, R.H.,(1990), Classical and Modern Regression with Applications , 2<sup>nd</sup> Edition , PWS-Kent Publishers , Boston.
39. Neter, J. et al ,(1990), Applied Linear Statistical Models : Regression , Analysis of Variance, and Experimental Designs , 3<sup>rd</sup> Edition, Irwin , Homewood , Il-60430 , Boston , MA02116.
40. Neter, J. et al ,(1993), Applied Statistics, 4<sup>th</sup> Edition, ALLYN and BACON- London.
41. Neyman, J. and Scott, E.L.,(1960), Gorrection for Bias Introduced by transformation of Variables , Ann. Math. Statist , 31, 643-655.
42. Pindyck, R.C. and RubinFeld, D.L. ,(1981), Econometric Models and Econometric Forecasts, 2<sup>nd</sup> Edition, New York , McGraw-hill.
43. Scheffe', H., (1953), A Method for Judging all Contrasts in the Analysis of Variance, Ann . Math. Statist., 40, 87-104.
44. Scheffe', H., (1959), The Analysis of Variance, Wiley, New York.
45. Seber, G.A.F.,(1977), Linear Regression Analysis , Wiley , New York.
46. Sen, A. and Srivastava, M.,(1990), Regression Analysis Theory, Methods and Applications ,Springer-Verlag, New York , Berlin, Heidelberg, Paris.
47. Stefansky, W.,(1971), Rejecting Outliers by Maximum Normed Residual , Ann. Math . Statist .,42, 35-45.
48. Stefansky, W.,(1972), Rejecting Outliers in Factorial Designs , Technometrics ,14, 469-479.
49. Walpole, R.E. et al ,(1998), probability and Statistics for Engineers and Scientists, 6<sup>th</sup> Edition - Prentice Hall International, Inc./Upper Saddle River, New Jersey , 07458.
50. Weisberg, S.,(1980), Applied Linear Regression , John Wiley , New York – Chichester , Brisbane, Toronto , Sfrngapore.

الملاحق

١. ملحق (١) جدول القيم الحرجة  $t_{\alpha}$  لتوزيع  $t$  .
٢. ملحق (٢) جدول القيم الحرجة  $f_{\alpha}(v_1, v_2)$  لتوزيع  $F$  عند  $(\alpha = 0.05)$  .
٣. ملحق (٣) جدول القيم الحرجة  $f_{\alpha}(v_1, v_2)$  لتوزيع  $F$  عند  $(\alpha = 0.01)$  .
٤. ملحق (٤) جدول القيم الحرجة لاختبار بونفروني .
٥. ملحق (٥) جدول المساحات تحت المنحى الطبيعي القياسي  $P(0 < Z < z)$  .
٦. ملحق (٦) جدول القيم الحرجة  $C_{\alpha}$  لاختبار الاعتدال .
٧. ملحق (٧) جدول معاملات كثيرات الحدود .
٨. ملحق (٨) جدول القيم الحرجة لدرين - واتسون .
٩. ملحق (٩) جدول القيم الحرجة  $\chi^2_{\alpha}(v)$  لتوزيع  $\chi^2(v)$  .

ملحق (١)

جدول القيم الحرجة  $t_{\alpha}$  لتوزيع  $t$

v	$\alpha$					
	.10	.05	.025	.01	.005	.001
1	3.078	6.314	12.706	31.821	63.657	318.31
2	1.886	2.920	4.303	6.965	9.925	22.326
3	1.638	2.353	3.182	4.541	5.841	10.213
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
60	1.296	1.671	2.000	2.390	2.660	3.232
120	1.289	1.658	1.980	2.358	2.617	3.160
$\infty$	1.282	1.645	1.960	2.326	2.576	3.090

المصدر : عن [Devore (1995)]



ملحق (٧)

جدول القيم الحرجة  $F_{\alpha}(v_1, v_2)$  لتوزيع  $F$  (عند  $\alpha = 0.05$ )

$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	
1	161.4	199.3	210.7	224.6	230.2	234.0	236.9	238.9	240.5	241.9	243.3	245.9	248.0	249.1	250.1	251.1	252.1	253.1	254.3
2	18.51	19.00	19.16	19.35	19.50	19.63	19.75	19.87	19.98	20.09	20.19	20.29	20.39	20.48	20.57	20.66	20.75	20.83	20.90
3	10.13	9.55	9.28	9.12	9.01	8.94	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.61	8.60	8.59	8.58	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.69	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.33	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.24	3.84	3.61	3.46	3.35	3.27	3.21	3.16	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.63	2.60	2.56
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.63	2.54	2.51	2.47	2.43	2.39	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.23	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.21	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.02
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.13	2.08	2.04	1.99	1.95	1.90	1.84
21	4.31	3.47	3.07	2.84	2.68	2.57	2.49	2.43	2.37	2.33	2.26	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.28	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.25	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.75
24	4.23	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.99	1.94	1.89	1.84	1.79	1.73
25	4.20	3.37	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.18	3.37	2.98	2.74	2.58	2.47	2.38	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.11	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.10	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.10	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.23	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.00	3.23	2.83	2.61	2.45	2.34	2.25	2.18	2.12	2.06	1.98	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.18	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
$\infty$	3.84	3.04	2.66	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.23	1.00

المصدر : عن [Devore (1995)]

جدول القيم الحرجة  $f_{\alpha}(v_1, v_2)$  لتوزيع F عند  $(\alpha = 0.01)$

		v1																		
v2		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	
1		100.0	99.90	99.80	99.70	99.60	99.50	99.40	99.30	99.20	99.10	99.00	98.90	98.80	98.70	98.60	98.50	98.40	98.30	
2		98.50	99.80	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.41	99.43	99.45	99.46	99.47	99.47	99.48	99.49	
3		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.50	26.31	26.12	26.22	26.13	
4		31.00	18.00	16.69	15.90	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.28	14.02	13.93	13.84	13.75	13.65	13.46	
5		16.26	13.27	13.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.71	9.65	9.47	9.38	9.29	9.20	9.11	
6		13.57	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	
7		12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	
8		11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	
9		10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	
10		10.01	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.83	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	
11		9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	
12		9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.46	
13		9.07	6.70	5.74	5.21	4.86	4.63	4.41	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.35	3.17	
14		8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	
15		8.68	6.36	5.42	4.89	4.56	4.33	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	
16		8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	
17		8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	
18		8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.67	
19		8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	
20		8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	
21		8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	
22		7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.41	
23		7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.36	
24		7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	
25		7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.17	
26		7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	
27		7.66	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.10	
28		7.64	5.45	4.57	4.07	3.75	3.53	3.37	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	
29		7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.24	2.05	
30		7.56	5.39	4.51	4.03	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.01	
40		7.51	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	1.94	1.80	
60		7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.35	2.20	2.12	2.03	1.94	1.73	1.60	
120		6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.64	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.55	1.38	
		6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.30	

ملحق (٤) جدول القيم الحرجة لاختبار بونفروني

$\alpha = 0.01$									
$v/m$	2	3	4	5	6	7	8	9	10
5	4.7733	5.2474	5.6042	5.8934	6.1384	6.3518	6.5414	6.7126	6.8688
6	4.3168	4.6979	4.9807	5.2076	5.3982	5.5632	5.7090	5.8399	5.9588
7	4.0293	4.3553	4.5946	4.7853	4.9445	5.0815	5.2022	5.3101	5.4079
8	3.8325	4.1224	4.3335	4.5008	4.6398	4.7590	4.8636	4.9570	5.0413
9	3.6897	3.9542	4.1458	4.2968	4.4219	4.5288	4.6224	4.7058	4.7809
10	3.5814	3.8273	4.0045	4.1437	4.2586	4.3567	4.4423	4.5184	4.5869
11	3.4966	3.7283	3.8945	4.0247	4.1319	4.2232	4.3028	4.3735	4.4370
12	3.4284	3.6489	3.8065	3.9296	4.0308	4.1169	4.1918	4.2582	4.3178
13	3.3725	3.5838	3.7345	3.8520	3.9484	4.0302	4.1013	4.1643	4.2208
14	3.3257	3.5296	3.6746	3.7874	3.8798	4.9582	4.0263	4.0865	4.1405
15	3.2860	3.4837	3.6239	3.7328	3.8220	3.8975	3.9630	4.0209	4.0728
16	3.2520	3.4443	3.5805	3.6862	3.7725	3.8456	3.9089	3.9649	4.0150
17	3.2224	3.4102	3.5429	3.6458	3.7297	3.8007	3.8623	3.9165	3.9651
18	3.1966	3.3804	3.5101	3.6105	3.6924	3.7616	3.8215	3.8744	3.9216
19	3.1737	3.3540	3.4812	3.5749	3.6595	3.7271	3.7857	3.8373	3.8834
20	3.1534	3.3306	3.4554	3.5518	3.6303	3.6966	3.7539	3.8044	3.8495
21	3.1352	3.3097	3.4325	3.5272	3.6043	3.6693	3.7255	3.7750	3.8193
22	3.1188	3.2909	3.4118	3.5050	3.5808	3.6448	3.7000	3.7487	3.7921
23	3.1040	3.2739	3.3931	3.4850	3.5597	3.6226	3.6770	3.7249	3.7676
24	3.0905	3.2584	3.3761	3.4668	3.5405	3.6025	3.6561	3.7033	3.7454
25	3.0782	3.2443	3.3606	3.4502	3.5230	3.5842	3.6371	3.6836	3.7251
30	3.0298	3.1888	3.2999	3.3852	3.4544	3.5125	3.5626	3.6067	3.6460
35	2.9960	3.1502	3.2577	3.3400	3.4068	3.4628	3.5110	3.5534	3.5911
40	2.9712	3.1218	3.2266	3.3069	3.3718	3.4263	3.4732	3.5143	3.5510
45	2.9521	3.1000	3.2028	3.2815	3.3451	3.3984	3.4442	3.4845	3.5203
50	2.9370	3.0828	3.1840	3.2614	3.3238	3.3763	3.4214	3.4609	3.4960
60	2.9146	3.0573	3.1562	3.2317	3.2927	3.3437	3.3876	3.4260	3.4602
70	2.8987	3.0393	3.1366	3.2108	3.2707	3.3208	3.3638	3.4015	3.4350
80	2.8870	3.0259	3.1220	3.1953	3.2543	3.3037	3.3462	3.3833	3.4163
90	2.8779	3.0156	3.1108	3.1833	3.2417	3.2906	3.3326	3.3693	3.4019
100	2.8707	3.0073	3.1018	3.1737	3.2317	3.2802	3.3218	3.3582	3.3905

تابع ملحق (٤)

$\alpha = 0.1$										
$\lambda/m$	2	3	4	5	6	7	8	9	10	
5	2.5706	2.9177	3.1634	3.3649	3.5341	3.6805	3.8100	3.9264	4.0322	
6	2.4469	2.7491	2.9687	3.1427	3.2875	3.4119	3.5212	3.6190	3.7074	
7	2.3646	2.6419	2.8412	2.9980	3.1276	3.2383	3.3353	3.4216	3.4995	
8	2.3060	2.5660	2.7515	2.8965	3.0158	3.1174	3.2060	3.2846	3.3554	
9	2.2622	2.5096	2.6850	2.8214	2.9333	3.0283	3.1109	3.1841	3.2498	
10	2.2281	2.4660	2.6338	2.7638	2.8701	2.9601	3.0382	3.1073	3.1693	
11	2.2010	2.4313	2.5931	2.7181	2.8200	2.9062	2.9809	3.0468	3.0158	
12	2.1788	2.4030	2.5600	2.6810	2.7795	2.8626	2.9345	2.9978	3.0545	
13	2.1604	2.3796	2.5326	2.6503	2.7459	2.8265	2.8961	2.9575	3.0123	
14	2.1448	2.3598	2.5096	2.6245	2.7178	2.7862	2.8640	2.9236	2.9768	
15	2.1314	2.3429	2.4899	2.6025	2.6937	2.7705	2.8366	2.8948	2.9467	
16	2.1199	2.3283	2.4729	2.5835	2.6730	2.7482	2.8131	2.8700	2.9208	
17	2.1098	2.3156	2.4581	2.5669	2.6550	2.7289	2.7925	2.8484	2.8882	
18	2.1009	2.3043	2.4450	2.5524	2.6391	2.7119	2.7745	2.8295	2.8784	
19	2.0930	2.2944	2.4334	2.5395	2.6251	2.6969	2.7586	2.8127	2.8609	
20	2.0860	2.2855	2.4231	2.5280	2.6126	2.6834	2.7444	2.7978	2.8453	
21	2.0796	2.2775	2.4138	2.5176	2.6013	2.6714	2.7316	2.7844	2.8314	
22	2.0739	2.2703	2.4055	2.5083	2.5912	2.6606	2.7201	2.7723	2.8188	
23	2.0687	2.2637	2.3979	2.4999	2.5820	2.6507	2.7097	2.7614	2.8073	
24	2.0639	2.2577	2.3909	2.4922	2.5736	2.6418	2.7002	2.7514	2.7969	
25	2.0595	2.2523	2.3846	2.4851	2.5660	2.6336	2.6916	2.7423	2.7874	
30	2.0423	2.2306	2.3596	2.4573	2.5357	2.6012	2.6574	2.7064	2.7500	
35	2.0301	2.2154	2.3420	2.4377	2.5145	2.5786	2.6334	2.6813	2.7238	
40	2.0211	2.2041	2.3289	2.4233	2.4989	2.5618	2.6157	2.6627	2.7045	
45	2.0141	2.1954	2.3189	2.4121	2.4868	2.5489	2.6021	2.6485	2.6696	
50	2.0086	2.1885	2.3109	2.4033	2.4772	2.5387	2.5913	2.6372	2.6778	
60	2.0003	2.1782	2.2990	2.3901	2.4630	2.5235	2.5752	2.6203	2.6603	
70	1.9944	2.1709	2.2906	2.3808	2.4529	2.5128	2.5639	2.6085	2.6479	
80	1.9901	2.1654	2.2844	2.3739	2.4454	2.5047	2.5554	2.5996	2.6387	
90	1.9867	2.1612	2.2795	2.3685	2.4395	2.4985	2.5489	2.5928	2.6316	
100	1.9840	2.1579	2.2757	2.3642	2.4349	2.4936	2.5437	2.5873	2.6209	

تابع ملحق (4)

$\alpha=0.05$									
w/m	2	3	4	5	6	7	8	9	10
5	3.1634	3.5341	3.8100	4.0322	4.2193	4.3818	4.5257	4.6553	4.7733
6	2.9687	3.2875	3.5212	3.7074	3.8630	3.9971	4.1152	4.2209	4.3168
7	2.8412	3.1276	3.3353	3.4995	3.6358	3.7527	3.8552	3.9467	4.0293
8	2.7515	3.0158	3.2060	3.3554	3.4789	3.5844	3.6766	3.7586	3.8325
9	2.6850	2.9333	3.1109	3.2498	3.3642	3.4616	3.5465	3.6219	3.6897
10	2.6338	2.8701	3.0382	3.1693	3.2768	3.3682	3.4477	3.5182	3.5814
11	2.5931	2.8200	2.9809	3.1058	3.2081	3.2949	3.3702	3.4368	3.4966
12	2.5600	2.7795	2.9345	3.0545	3.1527	3.2357	3.3078	3.3714	3.4284
13	2.5326	2.7459	2.8961	3.0123	3.1070	3.1871	3.2565	3.3177	3.3725
14	2.5096	2.7178	2.8640	2.9768	3.0688	3.1464	3.2135	3.2727	3.3257
15	2.4899	2.6937	2.8366	2.9467	3.0363	3.1118	3.1771	3.2346	3.2860
16	2.4729	2.6730	2.8131	2.9208	3.0083	3.0821	3.1458	3.2019	3.2520
17	2.4581	2.6550	2.7925	2.8982	2.9840	3.0563	3.1186	3.1735	3.2224
18	2.4450	2.6391	2.7745	2.8784	2.9627	3.0336	3.0948	3.1486	3.1966
19	2.4334	2.6251	2.7586	2.8609	2.9439	3.0136	3.0738	3.1266	3.1737
20	2.4231	2.6126	2.7444	2.8453	2.9271	2.9958	3.0550	3.1070	3.1534
21	2.4138	2.6013	2.7316	2.8314	2.9121	2.9799	3.0382	3.0895	3.1352
22	2.4055	2.5912	2.7201	2.8188	2.8985	2.9655	3.0231	3.0737	3.1188
23	2.3979	2.5820	2.7097	2.8073	2.8863	2.9525	3.0095	3.0595	3.1040
24	2.3909	2.5736	2.7002	2.7969	2.8751	2.9406	2.9970	3.0465	3.0905
25	2.3846	2.5660	2.6916	2.7874	2.8649	2.9298	2.9856	3.0346	3.0782
30	2.3586	2.5357	2.6574	2.7500	2.8247	2.8872	2.9409	2.9880	3.0298
35	2.3420	2.5145	2.6334	2.7238	2.7966	2.8575	2.9097	2.9554	2.9860
40	2.3289	2.4989	2.6157	2.7045	2.7759	2.8355	2.8867	2.9314	2.9712
45	2.3189	2.4868	2.6021	2.6896	2.7599	2.8187	2.8690	2.9130	2.9521
50	2.3109	2.4772	2.5913	2.6776	2.7473	2.8053	2.8550	2.8984	2.9370
60	2.2990	2.4630	2.5752	2.6603	2.7286	2.7855	2.8342	2.8768	2.9146
70	2.2906	2.4529	2.5639	2.6479	2.7153	2.7715	2.8195	2.8615	2.8987
80	2.2844	2.4454	2.5554	2.6387	2.7054	2.7610	2.8086	2.8502	2.8870
90	2.2795	2.4395	2.5489	2.6316	2.6978	2.7530	2.8002	2.8414	2.8779
100	2.2757	2.4349	2.5437	2.6259	2.6918	2.7466	2.7935	2.8344	2.8707

ملحق (٥)

جدول المساحات تحت المنحنى الطبيعي القياسي

$$P(0 < Z < z)$$

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

المصدر : عن [Daniel (1978)]

منحق(٦)  
جدول القيم الحرجة  $c_{\alpha}$  لاختبار الاعتدال

	$\alpha$		
	.1	.05	.01
5	.9033	.8804	.8320
10	.9347	.9180	.8804
15	.9506	.9383	.9110
20	.9600	.9503	.9290
n 25	.9662	.9582	.9408
30	.9707	.9639	.9490
40	.9767	.9715	.9597
50	.9807	.9764	.9664
60	.9835	.9799	.9710
75	.9865	.9835	.9757

ملحق (٧) جدول معاملات كثيرات الحدود

K	Polynomial	x=1	2	3	4	5	6	7	8	9	10	$\Sigma x^k$	$\frac{1}{k!}$
3	Linear	-1	0	1								2	1
	Quadratic	1	-2	1								6	3
4	Linear	-3	-1	1	3							20	2
	Quadratic	1	-1	-1	1							4	1
	Cubic	-1	3	-3	1							20	10/3
5	Linear	-2	-1	0	1	2						10	1
	Quadratic	2	-1	-2	-1	2						14	1
	Cubic	-1	2	0	-2	1						10	5/6
	Quartic	1	-4	6	-4	1						70	35/12
6	Linear	-5	-3	-1	1	3	5					70	2
	Quadratic	5	-1	-4	-4	-1	5					84	3/2
	Cubic	-5	7	4	-4	-7	5					180	5/3
	Quartic	1	-3	2	2	-3	1					28	7/12
7	Linear	-3	-2	-1	0	1	2	3				28	1
	Quadratic	5	0	-3	-4	-3	0	5				84	1
	Cubic	-1	1	1	0	-1	-1	1				6	1/6
	Quartic	3	-7	1	6	1	-7	3				154	7/12
8	Linear	-7	-5	-3	-1	1	3	5	7			168	2
	Quadratic	7	1	-3	-5	-5	-3	1	7			168	1
	Cubic	-7	5	7	3	-3	-7	-5	7			264	2/3
	Quartic	7	-13	-3	9	9	-3	-13	7			616	7/12
	Quintic	-7	23	-17	-15	15	17	-23	7			2184	7/10
9	Linear	-4	-3	-2	-1	0	1	2	3	4		60	1
	Quadratic	28	7	-8	-17	-20	-17	-8	7	28		2772	3
	Cubic	-14	7	13	9	0	-9	-13	-7	14		990	5/6
	Quartic	14	-21	-11	9	18	9	-11	-21	14		2002	7/12
	Quintic	-4	11	-4	-9	0	9	4	-11	4		468	3/20
10	Linear	-9	-7	-5	-3	-1	1	3	5	7	9	330	2
	Quadratic	6	2	-1	-3	-4	-4	-3	-1	2	6	132	1/2
	Cubic	-42	14	35	31	12	-12	-31	-35	-14	42	8580	5/3
	Quartic	18	-22	-17	3	18	18	3	-17	-22	18	2860	5/12
	Quintic	-6	14	-1	-11	-6	6	11	1	-14	6	780	1/10



ملحق (٧)

جدول القيم الحرجة لإحصاءات  $t$ -ن-والتسوية

Sample	Probability in Lower Tail Significance Level = $\alpha$	$k$ = Number of Regressors (Excluding the Intercept)									
		1		2		3		4		5	
		$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
15	.01	.81	1.07	.70	1.25	.59	1.46	.49	1.70	.39	1.96
	.025	.95	1.23	.83	1.40	.71	1.61	.59	1.84	.48	2.09
	.05	1.08	1.36	.95	1.54	.82	1.75	.69	1.97	.56	2.21
20	.01	.95	1.15	.86	1.27	.77	1.41	.63	1.57	.60	1.74
	.025	1.08	1.28	.99	1.41	.89	1.55	.79	1.70	.70	1.87
	.05	1.20	1.41	1.10	1.54	1.00	1.68	.90	1.83	.79	1.99
25	.01	1.05	1.21	.98	1.30	.90	1.41	.83	1.52	.75	1.65
	.025	1.13	1.34	1.10	1.43	1.02	1.54	.94	1.65	.86	1.77
	.05	1.20	1.45	1.21	1.55	1.12	1.66	1.04	1.77	.95	1.89
30	.01	1.13	1.26	1.07	1.34	1.01	1.42	.94	1.51	.88	1.61
	.025	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	.98	1.73
	.05	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
40	.01	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
	.025	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
	.05	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
50	.01	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
	.025	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
	.05	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
60	.01	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
	.025	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
	.05	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
80	.01	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
	.025	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
	.05	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
100	.01	1.52	1.56	1.50	1.58	1.48	1.60	1.45	1.63	1.44	1.65
	.025	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72
	.05	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

ملحق (٩)

جدول القيم الحرجة  $\chi^2_\alpha$  لتوزيع  $\chi^2$

v	$\alpha$									
	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.843	5.025	6.637	7.882
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.344	12.837
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.085	16.748
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.440	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.012	18.474	20.276
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.534	20.090	21.954
9	1.735	2.088	2.706	3.325	4.168	14.684	16.919	19.022	21.665	23.587
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.724	26.755
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.735	27.687	29.817
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.600	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.577	32.799
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.407	7.564	8.682	10.085	24.769	27.587	30.190	33.408	35.716
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.843	7.632	8.906	10.117	11.651	27.203	30.143	32.852	36.190	38.580
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.033	8.897	10.283	11.591	13.240	29.615	32.670	35.478	38.930	41.399
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.195	11.688	13.090	14.848	32.007	35.172	38.075	41.637	44.179
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.519	11.523	13.120	14.611	16.473	34.381	37.652	40.646	44.313	46.925
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.807	12.878	14.573	16.151	18.114	36.741	40.113	43.194	46.962	49.642
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.120	14.256	16.147	17.708	19.768	39.087	42.557	45.772	49.586	52.333
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.457	15.655	17.538	19.280	21.433	41.422	44.985	48.231	52.190	55.000
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.814	17.073	19.046	20.866	23.110	43.745	47.400	50.724	54.774	57.646
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.191	18.508	20.569	22.465	24.796	46.059	49.802	53.203	57.340	60.272
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.584	19.960	22.105	24.075	26.492	48.363	52.192	55.667	59.891	62.880
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.994	21.425	23.654	25.695	28.196	50.660	54.572	58.119	62.426	65.473
40	20.706	22.164	24.433	26.509	29.050	51.805	55.758	59.342	63.691	66.766

المصدر : عن [Devore(1995)]



## هذا الكتاب

يتناول هذا الكتاب كل ما يتعلق بتحليل الانحدار الخطي سواء البسيط أو المتعدد أو الانحدار الغير خطي . تستخدم تلك النماذج لغرضين على الأقل : عمل التنبؤات والحكم على قوة العلاقات . ولأن طرق الانحدار تمدها بالكيفية التي يتأثر بها متغير ما بمتغيرات أخرى فإنها أصبحت ضرورية في معظم مجالات الدراسة التي تشمل على العلوم الحيوية ، الفيزيائية ، العلوم الاجتماعية ، الصناعة ، الاقتصاد ، الطب ، ... الخ .

هذا الكتاب يصلح كمقرر لطلاب كثير من الكليات ، كما يصلح لأن يكون مقررًا لطلاب الدراسات العليا في جميع مجالات البحث العلمي . هذا ويمكن لطلاب الدراسات العليا في المجالات التطبيقية مثل الزراعة والطب والهندسة التركيز على الجانب التطبيقي من هذا الكتاب وتتبع حل الأمثلة .

يصلح هذا الكتاب أيضا لأن يكون مرجعا لأي باحث مع استشارة المتخصصين في الإحصاء وذلك لاختيار النموذج المناسب لتحليل البيانات . هذا ويمكن الاستعانة ببرامج الحاسب الآلي الخاصة بالانحدار لتنفيذ العمليات الحسابية مثل برنامج SAS أو SPSS أو Minitab ويفضل إجابة أكثر من برنامج حتى يمكن الاستفادة من إمكانيات كل برنامج .

المؤلف

Bibliotheca Alexandrina



0487719

مكتبة الأنجلو المصرية

THE ANGLO-EGYPTIAN BOOKSHOP

The World of Words & Thoughts

